EVOLUTIONARY TRAJECTORIES OF COGNITIVE ABILITIES AND OF THEIR PUTATIVE NEUROANATOMICAL AND ALLOMETRIC CORRELATES: TESTING NOVEL HYPOTHESES OF COGNITIVE EVOLUTION AND COGNITIVE INTEGRATION WITH PHYLOGENETIC COMPARATIVE METHODS

by

Heitor Barcellos Ferreira Fernandes

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment for the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2020

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Heitor Barcellos Ferreira Fernandes, titled "Evolutionary Trajectories of Cognitive Abilities and of their Putative Neuroanatomical and Allometric Correlates: Testing Novel Hypotheses of Cognitive Evolution and Cognitive Integration with Phylogenetic Comparative Methods" and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Aurelio José Figueredo — Date: 5/2/19

Melinda F. Davis — Date: May 2, 2019

William Jake Jacobs — Date: 2/5/2019

Horst Dieter Steklis — Date: 2 May 2019

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Aurelio José Figueredo — Date: 01/06/2020
Dissertation Committee Chair
Department of Psychology

Acknowledgments

I thank Mateo Peñaherrera-Aguirre and Michael A. Woodley of Menie for the valuable collaborations and their contributions, both theoretical and methodological, in this work and other, related research, past and ongoing. I also thank W. Jake Jacobs and Horst Dieter Steklis for their insight into the important role of the cerebellum in cognition, the empirical pursuit of which was fundamental to refining conclusions drawn from the analyses forwarded in the present work. Dieter and Netzin Steklis were also fundamental in shaping my always-updating notions of primate psychology and taxonomic diversity. Furthermore, I thank Melinda F. Davis for encouraging me to be methodologically innovative in this work. I must also thank Claudio S. Hutz for encouraging me to be ambitious when aiming for a PhD and in academic decisions. Finally, I thank Aurelio José Figueredo for being more than an advisor all the way: our intense collaboration and intellectual exchanges inspired this work and related professional investments.

Table of Contents

List of Tables

List of Figures

Abstract

The study of primate intelligence, and specifically of general intelligence, has progressed rapidly in the last two decades, however several issues remain unexplored. While neuroanatomical volume measures (NVMs) such as brain size, neocortex size, and absolute or relative size of other brain regions have been frequently framed as substrates for general intelligence, such claims are largely based on simple correlative analyses. Furthermore, while factor analytical techniques have identified a general factor among cognitive abilities when using datasets of species' mean performances, there have been no examinations of whether the common factor is predictably more strongly present in some primate lineages than in others, and whether relations among brain regions are similarly stronger in some primate lineages than in others. Here, such issues in the comparative literature on primate intelligence are addressed in a new set of studies that present two main novel contributions to the scientific understanding of primate intelligence:

First, the evolutionary patterns of the history of changes in general intelligence in primates is examined and compared to those behind the history of changes in brain size and the size of brain regions most commonly used in comparative cognition studies. Studying evolutionary patterns behind a trait permits examining how conserved it is across evolutionary time, how fast it has evolved, and the degree to which it has evolved in a particular direction (i.e., if natural selection regimes have been consistent). Phylogenetic comparative methods employed on datasets of primate species reveal that general intelligence has evolved at a faster pace than NVMs and it has evolved more consistently towards a selection optimum. In contrast to the overall emphasis given in the literature to brain size and neocortex ratio as substrates for

intelligence, the NVM with results that most strongly approached the patterns identified for general intelligence is residual cerebellar size (relative to body size).

Secondly, a hypothesis is advanced that species higher on general intelligence exhibit a stronger manifold (i.e., higher factor loadings) as general intelligence has previously been empirically associated with ecological generalism, rather than with specialism. As such, cognitive specialization and independence among abilities should be a hallmark of species that have not evolved strong general intelligence. The Continuous Parameter Estimation Model (CPEM) is used in a dataset of cognitive abilities in primate species, and largely confirms the hypothesis forwarded. However, when the same analytical approach is replicated using data on sizes of brain regions, it is found that brain size fails to predict or coevolve with factor loadings of brain regions. Similarly, telencephalon size (a brain structure that holds several of the regions theoretically proposed to serve as substrates for intelligence) fails to predict the strength of factor loadings of telencephalic regions.

This set of studies supports the notion that the comparability between the evolution of general intelligence and the evolution of volumes of neuroanatomical structures is more limited than previously thought. Alternative substrates for general intelligence are discussed.

Introduction

Scientific interest in the evolution of how cognitive abilities are organized or interrelated has seen a rise in nonhuman primate literatures. Importantly, the increased discussion has not been driven simply by the formulation of theories, simulations, or speculations, but has included empirical data and a multitude of analytical approaches at the cross-species level (Deaner et al., 2006; Fernandes et al., 2014; Reader & Laland, 2002; Reader et al., 2011), at the individual differences level (Shaw & Schmelz, 2017; Galsworthy et al., 2014; Hopkins, Russell, & Schaeffer, 2014), and at a mixed level (i.e., comparing individuals of different species; Herrmann et al., 2007; Woodley of Menie et al., 2017).

The notion that many cognitive abilities are not independent when primate species are compared, but rather grouped into somewhat broad domains (e.g., social intelligence, physical intelligence) is not novel or uncommon (Whiten & Byrne, 1988). Recently, however, hypotheses and examinations of the possibility of an existing common overarching factor among many abilities, including social and physical ones, have been advanced in primatology, both at the cross-species level (Deaner et al., 2006; Reader et al., 2011) and the individual differences level (Burkart et al., 2017). These publications involve studies both of laboratory tasks and of ethological observations of behavior, and follow decades of more anecdotal reports of the capacity for generalization of cognitive problem-solving across domains in many primate species. It is common to find descriptions of the capacity for object categorization and comprehension of generalized identity; categorization of abstract relations, and thus the capacity to interchangeably rely on different entities within a category to flexibly solve problems; insight into the affordances of tools or strategies without the use of a trial and error approach; understanding of object permanence; theory of mind and the capacity for deceit (countering the

expectations that an individual imagines that others have of it); and foresight in the form of anticipated problem solving or anticipated making of tools (indicative of some degree of planning). These capacities suggest that individuals of many species are capable of generalizing from one domain to another, to translate applications, and mentally manipulate information not only in parallel with physical manipulation of objects. Some researchers propose that the integrated set of such complex abilities is a "tool kit" with interdependences (Emery & Clayton, 2004), reflecting a *system* of cognitive abilities that arguably interact. Rigorous tests of general intelligence need to rely on more than these non-systematic observations however.

Findings of a general factor among performance scores on cognitive tasks when comparing individuals in a species indicate that individuals that perform well on one particular cognitive problem tend to perform well across others. This factor, frequently called Spearman's *g*, suggests that variations among individuals in performance are largely systematic across types of problems. For *g* to exist, some variation in performance must exist among individuals in the given species, even if it is small. The existence of a general factor across species, on the other hand, indicates that when species are compared in terms of their mean performance on several cognitive problems, those that have high average performance on a particular problem tend to perform well on average on others as well. This factor has been called *G*, to make it clear that a higher level of analysis is employed, rather than emerging from individual differences (Burkart et al., 2017; Fernandes et al, 2014). Studying *G* permits examining the *macroevolution* of cognitive abilities, whereas *g* more strongly reflects individual differences and thus microevolutionary processes.

While identifying *G* on a cross-species design suggests that there is a common latent factor behind multiple problem-solving cognition-related capacities that differentiates taxa, it

does not necessarily imply, by itself, that all taxa have an equally-cohesive *g* factor. It could very well be the case that, while most species in the dataset where *G* was identified are likely to exhibit a *g* factor, some may perform poorly on most tasks because perhaps they have specialized cognition for one or only a few narrow domains of problems. In other words, not necessarily all *g* factors are configured the same way (Arden & Zietsch, 2017), even if the identification of *G* in a given data set indicates that *most* species in it must, to a considerable extent, exhibit statistical integration (i.e., correlation) among the abilities tested. This is a research problem still to be resolved. Although all studies in the present work are concerned with *G*, Chapter 2 specifically will attempt to tackle this issue and examine if some primate taxa exhibit more weakly related cognitive abilities than others. While the study of *g* in nonhuman animals has progressed quickly with a multitude of studies demonstrating its highly replicable existence, heritability, and presence of additive genetic variance (Burkart et al., 2017; Galsworthy et al., 2014), comparatively few studies have focused on *G*, which thus necessitates further empirical attention.

The relatively recent laboratory studies systematically comparing primate species on cognitive tasks support the existence of the common, overarching factor *G* at the cross-species level (Deaner et al., 2006, Shultz & Dunbar, 2010). Taxa exhibiting high performance in certain tasks tend to do well across others as well. However, reliance solely on laboratory data to test the presence of a general factor of intelligence factor among species may have limitations, for several reasons: (a) laboratory tests can favor some species (i.e., being unfair to particular species), for example due to motor dexterity, visual acuity, or comfort in the setting; (b) laboratory tests may not reflect natural conditions (being thus ecologically invalid), and (c) data on laboratory tasks are available for a few species. For this reason, for example, Deaner and

colleagues (2006) had to avoid traditional factor analytic approaches and instead apply alternative statistical techniques simply to circumvent the amount of missing data and the little overlap among species in types of tasks on which they were tested.

As an alternative to this approach, Reader and colleagues (2011) collated counts of behaviors reflecting cognitive skill in the wild, from the existing primatological literature, to examine the existence of *G* in the primate order. Such data were made available for 69 primate species for five categories of behavior: tool use, extractive foraging, innovation, social learning, and tactical deception. Factor analyses using these five measures have confirmed the extraction of a common factor, explaining approximately 65% of their variance (Reader et al., 2011; further details are given in Chapters 1 and 2). The importance of this effect is compounded by the observation that this cross-species *G* factor is extremely robust to controls for a variety of potential confounds (e.g., phylogenetic relatedness, geographic proximity, group and population size, brain and body size, and exclusion of data from captive specimens). It is also compounded by the observation that measures that are more central to *G* (i.e., that exhibit stronger factor loadings) have evolved faster in primates and are less conserved across evolutionary time (Fernandes et al., 2014).

The factor analytic techniques employed in recent studies on cross-species cognitive data inform that *G* exists as a consistent and robust latent factor across primate species, but they do not reveal any more information about it. Its evolutionary history is unclear. Is the convergence of different abilities into the overarching factor present throughout primate history, or have correlations among abilities changed across evolutionary time? Has the evolution of *G* in primate phylogeny occurred in a particular direction (i.e., an optimal level)? And what are the biological substrates for the evolution of *G* in primates? Any evolutionary account of intelligence must

attempt to tackle these unexplored issues so that a clearer understanding of the nature of *G* is possible, beyond a description of its internal factor structure. Moreover, although intelligence studies in humans have been conducted for over a century as a prominent research area within psychology, these have not revealed information about the deep history of general intelligence, made possible through a comparative approach. Thus further comparative studies of general intelligence are required.

1. The Evolutionary History of *G*: What Is It, and Are the Most Commonly Studied Neuroanatomical Measures Convergent with It?

Phylogenetic comparative methods developed especially in the last two decades permit insight into the evolutionary trajectory of traits, not possible to be estimated before (Peñaherrera-Aguirre & Fernandes, 2018). A phylogeny is a representation of the evolutionary relatedness among groups of organisms; most commonly, the groups of organisms specified in phylogenies are species. A phylogeny of species contains information about which are sister lineages (i.e., having speciated more recently, such as chimpanzees and bonobos), and which are more distant (such as chimpanzees and ring-tailed lemurs), also informing *when* the lineages diverged. In other words, phylogenies contain information about how many million years ago a common ancestor for a given set of species gave rise to daughter lineages. Present knowledge on the phylogenetic tree of primate species is considerably precise, permitting comparative studies to employ it (Arnold et al., 2010; Perelman et al., 2011).

Just as a family tree is a useful tool for examining the history of a trait, a phylogenetic tree is useful for examining the evolutionary history of traits of interest. This is now feasible with data on extant species or populations, and relies on reconstructing the history of the traits in

question. With that approach, it is possible to examine *how* a trait has evolved (e.g., slow or fast; exhibiting conservatism or lability across time), and the mode of selection driving its evolution (e.g., directional and consistent selection across clades, or fluctuating; or rapid early in history and decelerating subsequently as niches were already filled), as will be presented in more detail in Chapter 1. Detailing the mathematical procedures behind such phylogenetic comparative methods is beyond the scope of this work, but knowing that they exist and what their affordances are leads to the realization that little has been done to understand the evolution of intelligence, even in primates – arguably the most frequently studied order in psychology. Using well-validated measures of cognitive abilities already demonstrated to converge into a common factor in primate species to examine the underlying evolutionary patterns and processes, is the primary goal of the present work. This will be addressed in Chapter 1.

Furthermore, it is possible to further understand a trait by understanding what it is related to (Cronbach & Meehl, 1955) and its biological substrates. Overwhelmingly, neuroanatomical volume measures (NVMs) are studied as substrates for intelligence and defended explicitly as proxies for it (Shultz & Dunbar, 2010). However, comparative studies have not produced a single, unified picture of the relationship between such measures and intelligence. Healy & Rowe (2006) summarized the state of the field as one of a bewildering array of correlations between NVMs and behavioral traits, a picture which shows little sign of resolving. This is because multiple measures are used (e.g., brain size, neocortex size, hippocampus size), through multiple operationalizations each (e.g., absolute size, size relative to the rest of the brain, or residualized against body size), and their correlations to intelligence are analyzed through several methods by different authors (e.g., bivariate correlations, multiple regressions; Barton, 2012; Deaner et al., 2000). Most striking is the overreliance simply on the correlational approach, as a

means to examine which NVM best represents intelligence. While the correlative approach does serve as one indication of the similarity in trajectories between two variables across evolutionary time, it should be seen exactly as simply an *indicator* of it. Complementary approaches are needed. This point is not made in the comparative literature. Take two hypothetical examples to illustrate this point:

(a) Let *i* and *j* be two brain regions, whose volumes for five species being analyzed are [1000, 2000, 3000, 4000, 5000] and [1001, 1002, 1003, 1004, 1005]. The correlation between *i* and *j* is objectively 1.0, but it is clear that changes in size across species in region *i* are immensely larger than changes in region *j*. They are thus evolving at overwhelmingly different rates in spite of a perfect correlation as measured by Pearson's *r* coefficient.

(b) Let *m* and *n* be two brain regions, with a correlation of .5 in terms of how their sizes vary across species. This magnitude is considered an important relation in the comparative literature on neuroanatomy and cognition and is more representative of relations actually observed than the 1.0 relation between *i* and *j* in example (a) (Deaner et al., 2000, 2007; Shultz & Dunbar, 2010). However, while the hypothetical region *m* exhibits consistent directional selection for increases in size, changes in region *n* somewhat resemble a random walk (i.e., with a much less consistent direction). Even though they share 25% of variance, the overwhelming 75% of specific variance in each of the two regions means that the selection regime behind them may be considerably different: even though part of changes in *m* are in the same direction as changes in *n*, a non-negligible number are in the *opposite* direction.

Compounding the limitations of the correlative approach, no consensus has been reached as to which measure and operationalization is the most correlated to intelligence (Barton, 2012; Deaner et al., 2000, 2007; Dunbar, 1992; Healy & Rowe, 2006; Shultz & Dunbar, 2010).

The present set of projects will examine the validity of the common approach of relying on NVMs as indicators or substrates of intelligence by testing the comparability of evolutionary trajectories of cognitive and neuroanatomical volume data in the primate phylogeny, and the comparability of the evolution of interrelations among components of general intelligence to the evolution of interrelations among neuroanatomical regions.

The last decade has seen publications with a large amount of data on NVMs, increasing both the number of species and the number of specimens and measurements per species for data on volume of the brain and its regions (Isler et al., 2008; Navarrete et al., 2018). This strongly complements the frequently employed data originally collected by Stephan and colleagues (1981), based on one or a few specimens for each primate species, and allows confident estimations of the comparability of evolutionary trajectories for NVMs so that they can be contrasted to the results for *G*. Still, scarcely have neuroanatomical data been examined for their evolutionary patterns with phylogenetic comparative methods – brain size in primates has been examined for its degree of phylogenetic conservatism, but nothing else (Kamilar & Cooper, 2013), while certain brain regions, such as cerebellum size, have been examined for their evolutionary rates only (Smaers et al., 2018). Analyses of both conservatism and evolutionary rates have been done on *G* (Fernandes et al., 2014). However, tests of fit for evolutionary models that indicate what selection regimes *G* and NVMs have been exposed to in primate phylogeny have simply not been conducted, let alone contrasted. Chapter 1 will thoroughly compare *G* and

the commonly studied NVMs in all of these aspects of evolutionary history: evolutionary conservatism, evolutionary rates, and selection regimes accounting for species differences.

2. Do All Species Need General Intelligence? Potential Variations In the Strength of the Manifold

Factor analysis reveals the *average* strength of correlations among measures on a given sample (Woodley of Menie et al., 2015). As already reviewed, studies that employed factor analytic techniques on measures of cognitive performance using data on primate species have revealed that, on average, abilities are integrated (i.e., they exhibit correlations and a single, common factor explains a large portion of their variance). However, it is possible that in some primate taxa, cognitive abilities are more integrated than in others: It is plausible that species low on *G* have simply not evolved a strongly integrated *G* factor – they have not required a domain-general cognitive system that recruits from more specialized abilities of more narrow domain. As such, species exhibiting low levels of *G* may perform well (and perhaps even excel) in some abilities independently of others, rather than exhibiting consistency in performance across tasks. While consistency across tasks implies strong factor loadings for most or all cognitive tasks, species with high performance in few tasks would show more unique, task-specific statistical variance. The novel statistical approach of continuous parameter estimation (Gorsuch, 2005) contains affordances that make feasible such an examination of variations in factor loadings in a dataset of species, as will be proposed and implemented in Chapter 2.

Examining this novel hypothesis, like examining the evolutionary trajectories of *G* compared to those of NVMs, illuminates what is known about the nature of *G* by going beyond a simple description of its average factor pattern. A hypothesis that clades differ in how related

cognitive abilities are, as fully outlined and unpacked in Chapter 2, derives from and requires evolutionary reasoning about the purpose of *G.* Species facing adaptive problems or opportunities that require *G* should not only evolve higher *levels* of it, but more *generality* of performance (i.e., successful problem-solving for multiple diverse challenges). As such, if this hypothesis is supported, it can be concluded that *G* is aptly termed *general* intelligence.

Just as the approach of looking at the comparability between *G* and NVMs through the lenses of phylogenetic comparative methods (rather than again through correlations) permits a novel perspective on that research problem, examining whether *G* exhibits factor loadings that are stronger in some taxa than in others permits a novel solution to the problem of whether *G* is a widespread phenomenon. The debate as to the existence of *G* has relied mostly and repeatedly on traditional factor analyses (Burkart et al., 2017); the approach proposed here permits testing the balanced notion that *G* is widespread in those species that need *G* for their local problem-solving and thus for their adaptation; in others, a manifold need not exist. In other words, whether factor loadings reveal a common factor in some taxa, but not in others, can be tested in a *systematic and predictable* manner.

3. Are There Equivalent Variations in the Strength of the Manifold among Brain Regions? Further Tests of the Comparability between *G* and the Neuroanatomical Volumetric Approach

Considering the overwhelming weight given to NVMs in the comparative literature on cognition so far, further tests of the similarity of their evolutionary history to that of cognitive abilities are needed. In keeping with the goal of contrasting the evolutionary trajectory of *G* and of NVMs, it is important to realize that the relationship between (a) brain size and (b) the

strength of correlations among sizes of brain structures can also be tested. This parallels the tests done in Chapter 2, and permits examining if there is any systematic and predictable variation in how strongly brain structures are associated in terms of their size.

Studying this also has the potential of elucidating why some researchers conclude that brain structures have varied in size in concert through time whereas others argue for independent or so-called mosaic evolution of brain regions (Finlay et al., 2001; Smaers & Soligo, 2013). Like the debate about which NVM best predicts intelligence, the debate between those that defend mosaic evolution of brain regions and those that argue for concerted changes in sizes of brain regions has failed to examine potential reasons for variation among clades in extent of concerted or mosaic brain evolution. It is possible that in some clades, mosaic evolution is more strongly the norm, whereas in others the converse is true (Barton & Harvey, 2000). Larger brains, if indeed reflective of intelligence, may necessitate more coordination among structures, perhaps leading to concerted evolution of brain regions. This will be tested in Chapter 3.

4. General Significance

Because tests of hypotheses implemented in the following studies are novel both in theory and in statistical application of relatively recent approaches not yet used in the *G* literature, caution will be used by carefully examining potential confounds and alternative analytical designs[1]. Together, these three studies have the potential of revealing signs of adaptative value in increasing levels of *G* in primate history, increasing strength of its manifold, and its comparability to so often used NVMs. Such a comparison is essential, considering the

[1] However, so as to not have the reader distracted from the main objectives of each study, several of these complementary analytical approaches will be presented in the Appendices.

weight that volumetric measures for the brain and some brain regions have carried for decades as the golden standard of substrates of intelligence in primate studies, leaving other putative anatomical substrates of intelligence (e.g., myelination, neuron density, short- and long-range connectivity, and even gyrification; Dicke & Roth, 2016; Roth & Dicke, 2012) with little attention, and perhaps consequently, little available comparative data. Already in 1979 (p. 85), noticing how simplistic or limited this approach might be, Ralph Holloway expressed the concern that:

> "By cathecting on size alone, all evolutionary paradigms become reduced to natural or genetic selection operating on incremental size increases and behavioral efficiency, which always has the underlying implicit structural argument that 'intelligence' equals 'brain size.'"

The studies in the present work assess if future steps in studies of the neuropsychology of intelligence could benefit from continued exploration of the volume measures as substrates for *G,* or if further attention should be given to alternative characteristics of the brain. Together, they also have the potential of informing the degree to which *G* exhibits signs of adaptive value, and how much it varies predictably across species not only in *size*, but also in *internal structure*. By the same token, they have the potential of informing if associations among sizes of brain regions vary predictably across primate species. These insights can help guide future directions for the study of general intelligence and NVMs.

Chapter 1. Macroevolutionary Patterns and Processes for General Intelligence (*G*) and for Commonly Used Neuroanatomical Volume Measures in Primates: Low Convergence Indicates Largely Independent Selection Regimes

1. Introduction

Measures of cognitive performance are found to be highly associated in comparative analyses of primate species, giving rise to a common factor of intelligence *G*. The *G* factor is found both for ethological counts of novel problem-solving abilities (as measured, for instance, through the observation of innovation rates, extractive foraging, tool use, tactical deception, and social learning; Reader et al., 2011) and for controlled, laboratory tasks (Deaner et al., 2007). In fact, these methods lead to correlated indices of general intelligence at the cross-species level (Reader et al., 2011). These findings replicate evidence of a *g* factor across many mammal species studied so far, including primates and also rodents (for reviews, see Burkart et al., 2017; Galsworthy et al., 2014; Shaw & Schmelz, 2017).

Importantly, it appears that measures on which *G* loads more strongly have been a main focus of selection pressures throughout primate evolutionary history, more so than the more specialized abilities, as the former exhibit faster evolutionary rates and more lability (Fernandes et al., 2014). In other words, *G*-central indicators display less evolutionary conservatism from ancestor to daughter lineages, with closely-related species tending to have more rapidly diverged with respect to more *G*-loaded measures of performance. These findings suggest that primate species with high success in solving a particular task also tend to exhibit high success in other cognitive adaptive challenges, and that a general factor is thus more parsimonious than

explanations involving the evolution of distinct and specialized abilities (for a review, see Burkart et al., 2017).

In spite of the increased interest and largely consistent findings on general intelligence across primates, there has been little empirical exploration of the neurological bases of general intelligence apart from correlative analyses with volumetric measures. The 'folk impression' - that brain size or the size of certain regions reflects intelligence (and thus can be used to track this across evolutionary history) - has been fundamental in the research traditions of anthropology and, more specifically, primatology. This intuitive impression has been a component of evolutionary thinking since its early history – Darwin (1871, p. 145), for instance, presumed that no one would doubt "that the large size of the brain in man, relatively to his body with that of the gorilla or orang, is closely connected with his higher mental powers". This notion has also permeated influential works in zoology throughout the 20th century (e.g., Jerison, 1973). Many authors suggest that neuroanatomical volume measures (NVMs) can and should be used as a "proxy for intelligence" at the cross-species level (Shultz & Dunbar, 2010, p. 259). A long-lasting debate exists about *which* NVM (i.e., the whole brain, the neocortex, the hippocampus, etc.) is mainly responsible for intelligence, while assuming that at least one of these measures is to a large extent responsible for the evolutionary trajectory of intelligence across the primate order. Each measure appears to exhibit associations with cognitive performance at the cross-species level, but it is also essential to understand the limitations so that a comprehensive comparison can be made and questions that help us move forward can be better framed. A review of the information available regarding each of the most commonly used and defended NVMs, and of their relations to *G*, is conducted below, before the novel phylogenetic analyses on them and on *G* are detailed and compared as an original contribution of the present Chapter.

1.1. Brain Size

Empirical evidence supports the view that absolute brain size predicts cognitive ability in comparative studies of mammals (e.g., Barrickman et al., 2008; Byrne & Corp 2004; Deaner et al. 2007; Lefebvre et al., 2004; Reader et al., 2011), and specifically in primates it shows higher evolutionary correlations with *G* than other commonly used NVMs, such as the residual of brain volume against body size, neocortex volume and neocortex ratio (i.e., the ratio of neocortex volume relative to the volume of the rest of the brain), among others (Deaner et al., 2007). Absolute brain size appears also to be predictive of related traits, such as problem-solving tasks requiring self-control (MacLean et al., 2014). Species differences in proxies for broader intelligence tests, such as the transfer index test, are also predicted by absolute brain size (e.g., Gibson et al., 2001). As bigger brains can contain more neurons (and brains of larger size than expected for a given body size may contain what is commonly called *extra neurons*; Jerison, 1973), the rationale is simply that bigger brains must have more processing capacity. Moreover, as *G* by definition is a complex and domain-general trait, it is not expected to be highly localized, but to draw from networks involving many brain regions, justifying the interest in total brain size as a substrate for *G*. Considering the strong allometric relations between body and brain size, and between brain size and the size of specific regions of the brain, many researchers use relative, residualized volume measures as an indicator of intelligence (for a review, see Healy & Rowe, 2007). These approaches have largely replaced the early reliance on the encephalization quotient (for a recent review, see Peñaherrera-Aguirre et al., 2017) which is highly unreliable as its equation varies strongly depending on which species are added and which are removed from the model.

Brain size is clearly also used due to the principle of parsimony. As it has a considerable correlation with the size of most brain structures, and a non-negligible correlation with several non-volume measures that may affect cognition, such as gyrification, it is argued to serve as a good catch-all measure to explain intelligence (Falk & Gibson, 2001).

However, approaches relying on brain size, be it absolute or relative to body mass, are not without criticisms and there is evidence that makes its defense as the best neuroanatomical measure for *G* dubious at best. Apes, including humans, and monkeys do not exhibit the largest brain, either in absolute terms or relative to body size (Dicke & Roth, 2016), contradicting the contention that this is a good neuroanatomical measure of intelligence. Furthermore, different innovation rates and problem-solving capacities are observed in species with similar brain sizes (Forss et al., 2016; Navarrete & Laland, 2015), indicating that other substrates for cognitive performance need to be invoked. These criticisms of the brain size approach are not new, however. Several scholars have expressed concerns with the recent rise of interest in absolute brain size in the field of complex cognition (e.g., Chittka & Niven, 2009; Healy & Rowe, 2007), renewing doubts and criticisms expressed decades ago (e.g., Holloway, 1966a,b), and considering that the field was developing in the direction of analyzing more fine-grained neuroanatomical candidates for intelligence rather than examining a broad and unspecific proxy such as brain size (Healy & Rowe, 2007).

1.2. Major Candidate Brain Structures

1.2.1 Neocortex size

The neocortex has been proposed by many researchers to be responsible for complex cognitive information processing, especially in the form of the social brain hypothesis (Barton,

1996; Dunbar, 1992; Dunbar & Shultz, 2007; Shultz & Dunbar, 2010). This hypothesis proposes that, in order to cognitively monitor social interactions in increasingly complex groups (in terms of size and forms of relations) in primate evolution, larger neocortices were necessary. As such, it is a hypothesis about cognitive processing and executive functions (Shultz & Dunbar, 2010), rather than necessarily specifically about general intelligence. Still, the neocortex has been compared to other volume measures in examinations of the best neuroanatomical predictors of intelligence (e.g., Deaner et al., 2007), and also to ethological counts of intelligence-related behaviors (social learning, tool use, and innovation rates; Reader & Laland, 2002). However, neocortical size also correlates importantly with rates of tactical deception at a cross-species level in primates (Byrne, 1996), a variable that does appear to be an integral part of the *G* nexus (Fernandes et al., 2014; Reader et al., 2011).

The case for the neocortex as a neuroanatomical indicator of intelligence, is further made by Kaas and Herculano-Houzel (2017), who suggest that, given the positive relation between neocortical volume and number of neurons, a larger neocortex has more neurons to analyze sensory inputs. Furthermore, the expanded cortical sheet would contain more cortical areas, permitting more computationally intensive information processing and storage, and decision making. However, they argue that this should manifest as cognitive specialization (Kaas & Herculano-Houzel, 2017), whereas general intelligence is, by definition, generalized contrary to this expectation, reducing the plausibility of the hypothesis of cognitive evolution focused on cortical complexification as a candidate explanation.

The most common use of neocortex volume as a proxy for intelligence is the neocortex ratio (Dunbar, 1992; Shultz & Dunbar, 2010), operationalized as the size of the neocortex divided by the size of the rest of the brain. However, it too is not without criticism: It has been

questioned why the neocortex ratio would be a proper statistical approach to understanding the evolution of the neocortex, as the enlargement of any other parts of the brain would decrease the neocortex ratio, leading to expectations of lower intelligence. It is unclear why enlargements in other areas, several of which are known to be involved in complex cognition and to be part of neural circuits that include the neocortex, should be interpreted as decreasing intelligence (Gibson et al., 2001). Neocortical volume residualized against body size has also been used as an approach occasionally, although more in the sociality literature than in cognition studies (Deaner et al., 2000).

Among components of the neocortex, the frontal lobe has been central to much discussion especially for the evolution of human intelligence. A common assumption is that high relative enlargement of the frontal lobe is the hallmark of human brain evolution. However, recent evidence using correctly scaled measures and phylogenetic approaches indicates that no such relative enlargement has occurred (Barton & Venditti, 2013), either for humans or apes in general. In fact, several non-hominoid branches in the primate phylogeny exhibit faster evolutionary rates than those for the former taxa.

1.2.2. Cerebellum size

More rarely discussed, but still tested and highlighted especially in more recent publications, is the potential role of the cerebellum in intelligence. It has been argued that an excessive emphasis on neocortical volume has obscured the putative role of the cerebellum and led to its relative neglect (Barton, 2012). Several lines of research indicate that cognitive capacities are predicted by cerebellar size, be it its absolute size or relative to broader measures such as body size. Firstly, as cerebellum size increases in primate lineages, its neuron density

exhibits a much less noticeable decline compared to the neuron density decline in larger neocortices (Barton, 2012). Secondly, the cerebellum has undergone rapid evolutionary expansion in the great ape clade (which also exhibits high *G*; Reader et al., 2011) (see Miller et al., 2019, for a brief review). Along with such volume increases, there are more cerebellar-neocortical connections in apes (Barton, 2012; Rilling, 2006), with these two structures possibly evolving as a coordinated system (Barton & Harvey, 2000). While the cerebellum is usually considered to mainly have a role in motor control, it has long been proposed that it actually is a modulator and augmenter of neurologic function: Connections to motor areas would increase the skill of movement, while connections to cognitive areas would improve the skill of thought-intensive problem solving (Leiner et al., 1989; Snider, 1950). Thus the observed increased connectivity to neocortical areas may be responsible for increased general cognitive ability.

Evidence has accumulated that the cerebellum is involved in many cognitive domains, including planning and decision-making, associative learning, working memory, spatial and episodic memory, mental rehearsal, event prediction, and imitation (for a review, see Barton, 2012). In fact, cerebellar size (residualized against body size and brain volume) is more predictive than neocortical size of tool use and extractive foraging (measures implicated in general intelligence; Reader et al., 2011; Fernandes et al., 2014) (Barton, 2012). More recent evidence suggests that evolutionary increases in cerebellar size, especially in the lateral cerebellar hemispheres, are correlated with general intelligence in primates through multiple independent evolutionary occurrences (Smaers et al., 2018). Considering these lines of evidence altogether, cerebellar size must be considered as one of the main and increasingly studied candidates in terms of the volumetric substrates of *G*.

1.2.3. Hippocampus size

Although also rarely proposed as being directly responsible for broad, general intelligence, the hippocampus is often invoked as a structure that is integral to the information maintenance and cognitive control functions of the neocortex, especially the pre-frontal cortex (Blair, 2006), and as such figures as a candidate region for neuroanatomical regions responsible for executive functions (e.g., memory and attention) and intelligence when its absolute size is used to predict these variables (Shultz & Dunbar, 2010). Although positive associations are found with executive functions, little has been explored about its relationship with *G* as it is uncommon to examine hippocampus size in comparative analyses of intelligence, especially considering the limited amount of data on hippocampal volume for primates compared to other measures, and considering that the differences among primate species in hippocampal size are slight (Stephan et al., 1981).

1.3. Beyond Correlations

That volumetric measures of the brain, the neocortex, and the hippocampus all show correlations with intelligence is little debated. However, examining the evolutionary associations among traits, and therefore testing whether one may function as the main factor for variation in another during evolution is a more complex endeavor than simply looking at the correlations between variables. Inferring evolutionary processes for traits from correlations across extant species can be misleading. A trait that exhibits strong correlation with another may be under a different selection regime, display a different evolutionary trajectory, and only constrain the evolution of the other trait (thus permitting a window of variability within which no evolutionary influence may be exerted), rather than function as a *driver* of its evolution. This case can be

illustrated with recent studies that have identified different evolutionary trajectories for brain and body size in spite of strong correlation: Analyses of cichlid adaptive radiation indicates that body size exhibited recent bursts of rapid evolution that were not found for brain size – the latter evolved in a gradual manner (Gonzalez-Voyer et al., 2009). Many selection mechanisms may be responsible for the relative evolutionary independence of phenotypically and genetically related traits, among which it has been found, in a study of pinnipeds, that body and brain size evolutionary trajectories may be decoupled by sexual selection (Fitzpatrick et al., 2012). Other, hypothetical, scenarios are also possible: increases in brain size in a lineage may require body sizes that accommodate them, but larger body size may have already evolved before due to predation risk or other selection pressures. A similar rationale may be applied to the association between intelligence and NVMs: it is not necessarily the case that they need to evolve in tandem. *G* and certain (or all) NVMs may have been under different selection regimes and thus may exhibit different evolutionary trajectories, in spite of correlations.

Another issue with the volumetric approach to understanding intelligence is that comparative studies indicate that neuronal density and gray matter density in many structures of the brain tend to be smaller in species with a larger brain volume (Barton, 2006; which also applies to the frontal lobes; Semendeferi et al., 2011). As such, evolutionary increases in NVMs can be deceptive: For example, in apes the cerebral cortex represents 70-82% of brain mass (more than in other primates) but holds only 19-30% of brain neurons (similar to or less than other mammals; Herculano-Houzel, Collins, Wong, & Kaas, 2007). Consequently, it is an expansion of white matter that is favored in larger brains to maintain conduction speed (Barton, 2006; Herculano-Houzel et al., 2010; Wen & Chklovskii, 2005), with processing power increases thus not being the main outcome of the evolution of larger brains. While increased connectivity

between closely positioned neurons may exist in larger brains, it is decreased among different regions of the brain in spite of a higher number of axons in the white matter (Semendeferi et al., 2011). As general intelligence is influenced by cortical connectivity, it is likely to have considerable independence from brain or neocortex volume, being also importantly influenced by other, non-volumetric factors.

It is undeniable that NVMs and intelligence show correlation at the cross-species level. However, to further understand the associations of *G* and NMs and examine the degree to which they share an evolutionary history, it is necessary to compare their evolutionary *processes*, namely what selection regimes they have been under, and to test if they are convergent. It is also essential to compare their rates of evolution, because even though they may be evolving in the same directions (with one increasing when the other increases, and decreasing when the other decreases, and thus are positively correlated across evolutionary history), *G* might be evolving at a faster rate than NVMs, thus necessitating that other covariates be invoked as potential substrates. The present study aims to address these questions.

In sum, this study has the goal of examining the evolutionary history of *G* in primates as compared to that of the above commonly employed and defended neuroanatomical volume measures. Exploring all possible brain areas, each operationalized in many ways (e.g., absolute size, residualized against body size, or using its ratio to the rest of the brain, etc.) would constitute a largely exploratory approach that invites capitalization on chance and makes theory construction difficult. Rather, only already used NVMs and their specific operationalizations that have led to positive correlation coefficients with intelligence measures will be examined. Thus, this study builds on previous hypotheses and evidence.

2. Method

2.1. Measures

2.1.1. G

Data on *G* were compiled by Reader and colleagues (2011), covering over 4000 publications, and 69 species including four ethological dimensions or classes of cognitive abilities (social learning, tool use, innovation, and extractive foraging). Observed instances of a five dimension (tactical deception) were obtained from Byrne and Whiten (1988). These five variables refer to ethological counts of behaviors recorded in the literature (i.e., the primary data collection). Ethological counts for each of the five categories of behavior were registered in the database for each species (i.e., the secondary data collection), and residualized against research effort for the respective species (i.e., the number of ethological works published in general for each species across literatures; Reader et al., 2011).

To permit an understanding of the meaning of *G*, it is important to conceptualize the five cognitive abilities comprised in it:

(i) Tool use: The employment of an object to alter the form, position, or condition of another object, organism, or the user itself. This is a classic intelligence measure in primatology, and in anthropological studies of human evolution (see Matsuzawa, 2001; McGrew, 1993) as performance on this domain is considered a direct indicator of the capacity to control one's immediate environment (Darwin, 1871; Gibson and Ingold, 1993; Washburn, 1959; Wynn, 1988). Tool use has been demonstrated to be associated with other indices of intelligence in primates and other animals (Lefebvre et al., 2004; Reader & Laland, 2002).

(ii) Extractive foraging: The capacity to identify, access, and therefore obtain food items that are concealed in some way (e.g., buried, covered, or encased within a shell). This capacity relates to general intelligence both theoretically and empirically (Gibson, 1986, Parker and Gibson, 1977, Reader et al., 2011; van Schaik and Isler, 2012), and exhibits considerably high evolutionary lability and evolutionary rates among primate species (Fernandes et al., 2014).

(iii) Innovation: A new solution to complex social or technical problem, or the application of an already-existing solution to a novel complex problem. It relates to the capacity to ontogenetically adapt to new environmental conditions, and is thus considered a proxy for intelligence (Lefebvre et al., 2004; Reader & Laland, 2002; Sol et al., 2005, 2008).

(iv) Social learning: The capacity to learn skills and acquire information from family members and others in the social group. Social learning is often considered a central component of social or Machiavellian intelligence (Byrne & Whiten, 1988, Humphrey, 1976; Whiten and Byrne, 1997) and associated with broader intelligence (Reader et al., 2011).

(v) Tactical deception: Behaviors that attract the attention of others to or away from self to confuse, mislead, manipulate, exploit, take advantage of, or otherwise damage others to one's benefit in a way that gives the appearance of non-malevolence (Byrne & Whiten, 1985). This capacity is also central do Machiavellian intelligence. In spite of exhibiting the lowest factor loadings from *G* and lowest evolutionary rates among these five ethological measures of cognitive ability (Fernandes et al., 2014), its evolutionary lability is higher than that of most mammalian traits, even behavioral

(Blomberg et al., 2003) and its factor loading from the *G* factor is moderate to high (Reader et al., 2011).

For a much expanded conceptualization, a review of existing operationalizations used in the original ethological observations compiled by Reader and colleagues (2011) for the databased used in the present study, and behavioral examples of the five cognitive abilities, see Appendix A.

Primary Data Collection

Reader et al.'s (2011) database includes publications from a wide span of years, with thus primary data collection having been implemented at different moments of the development (and refinement) of ethological methods. Many initial observations recorded (although certainly a few recent ones as well; see Appendix A for examples) derive from unstructured field observation and are anecdotal in nature. Such pilot observations, however, were essential for the development of basic understandings of primate behavior and the subsequent organization of detailed, rigorous ethograms. Ethograms, progressively more used across time in the primary data contained in the database employed here, are inventories of behaviors organized taxonomically; that is, there is commonly a hierarchical categorization of behaviors (Herring, 2017), ranging from very specific behaviors (e.g., visually following the gaze of kin), to broad categories of interrelated behaviors (e.g., social behavior, including the above example, but also play behaviors, affiliative behaviors, etc.). Though ethograms may vary from research group to research group based on their particular focal interest, definitions included in ethograms tend to be based on previous observations, discussions, and agreements available in the literature.

Through the refinement of ethograms, the purpose of clear, pre-established definitions was noted: Many authors initially did not carefully provide an explicit definition of their behaviors of focus in their reports, analyses, and discussion (Shumaker et al., 2011). Note that careful application of the detailed definitions reviewed in Appendix A, in the ethological observations published throughout the primatological literature and collated by Reader et al. (2011), were left to the primatologists who submitted the records, and in some cases their scope was probably stretched or reduced. Careful adherence to strict and detailed definitions increased as ethological practices were refined and inter-observer reliability sought.

Importantly, the field of ethology, including its use in primatology, has evolved to rely on multiple observers, and to prioritize records exhibiting high inter-observer reliability. Though the estimation of reliability is still more common in studies relying on a laboratory setting (e.g., Povinelli et al., 2009) and personality assessment (Briffa & Weiss, 2010), it grew over time as an important aspect of all ethological observations. Studies have been developed exclusively to test inter-rater reliability of ethogram-guided observations in the wild, demonstrating that, with proper training and ethogram clarity, very high inter-observer reliability estimates are attained (e.g., Coelho & Bramblett, 1981).

Many methods for behavioral measurement have been developed, for diverse and complementary, but sometimes partly overlapping purposes (MacDonald, Stewart, Stopka, & Yamaguchi, 2000). Considering the number of publications included in the database compiled by Reader and colleagues (2011) and their spread across decades, these alternative approaches have each been used numerous times in the database. As such, methodological approaches for primary data collection in studies collated in the database used in the present study were not homogeneous. Sampling methods include:

a) Ad libitum sampling, where behaviors are recorded in an informal, non-systematic way but as much as possible is recorded. This approach is most useful for pilot observations to obtain a survey of potential behaviors of future focus;

b) Focal sampling, where attention is given to one individual specifically but to all its behaviors, and its observations may or may not be systematically guided by an ethogram. This approach is especially useful if the researcher is interested in long, complex behavior that cannot be easily identified if attention is diffused across the community or sparsed across time;

c) All-occurrance sampling, where attention is given to a specific type of behavior, irrespective of which individual is performing it, providing information on the natural rate of occurrance of a behavior of interest;

Irrespective of the approach employed, data sampled can also be *recorded* in a variety of ways, employed to varying degrees throughout the literature used as source for the database used in the present study:

a) Continuous: The duration, frequency, and sequence of all behaviors are recorded;

b) Instantaneous: Current behavior(s) identified in a predetermined interval (e.g., every 30 seconds) are recorded and detailed;

c) One-zero: Current behavior(s) identified in a predetermined interval are only marked as eithered having happened or not, with not detailed descriptions.

Certain combinations predominate: focal sampling and continuous record; scan sampling and instantaneous record; and focal sampling and one–zero record. Instantaneous and one-zero approaches can be useful to obtain a general idea of the distribution of behavior states in a

community, being most useful if potentially observable behaviors of interest are easily identifiable and of relatively long duration, otherwise they can be missed. Simultaneous and continuous focal sampling of all observable individuals in a group is arguably the most accurate and information-rich combination of approaches, but it is not commonly attainable, and unlikely that most studies in the database used in the present study employed this combination (MacDonald et al., 2000). Computer-based recording and the use of software to subsequently track and identify individuals and behaviors has facilitated the implementation of simultaneous and conitnuous focal sampling of as many individuals as possible, but it is both recent (thus not being readily widespread in most of the past accumulated literature and therefore the present database) and frequently invasive to the community, compared to alternative but more limited technologies (e.g., radiotracking) or compared to a single-researcher observation in certain circumstances that might alert or arouse individuals observed. Moreover, though continuous focal sampling can be information-rich, systematic comparisons of sampling and recording approaches have shown that instantaneous (i.e., interval) recording can in certain situations be not only more efficient but also reduce bias, as it is easier to employ in difficult conditions and may reduce researchers' bias toward certain forms of behaviors (Rose, 2000).

Secondary data collection

Judgment of whether a given behavior pattern sampled, recorded, and published qualified in one of the five categories of behavior was made the expert author of each original article (Reader et al., 2011). For instance, behaviors originally described as “novel”, “innovative”, “creative”, “never seen before”, among close synonyms, were included in the “innovation” category. Similarly, behavior descriptions containing keywords that connoted that an individual

learned about resources, the environment, or others from conspecifics (rather than by themselves) were included in the "social learning" category. As previously argued by Lefebvre (1997) and Reader and colleagues (2011; Reader, 2003; Reader & Laland, 2001), this approach avoids subjective bias on the part of the researcher responsible for collating data for the database. It must be emphasized that the behavior reports were not derived from uninformed casual observations, but rather came from experienced, established scientists extensively familiar with their subjects (Reader, 2003). Furthermore, observers themselves have more information about the context in which behaviors occurred than Reader and colleagues had for the collation of the database, and the context matters for the categorization of the behavior: for instance, the act of placing a potato inside a small body of water (as observed in Japanese macaques; Kawamura, 1959) may be indicative of innovation, of tool use, or of deception, depending on the context in which it happens, and the particular purpose it serves in that specific behavior instance. If it is a novel occurrence, it may be classified as innovation; if instead the behavior is common in the species (as it is today in Japanese macaques) but the behavior in the case of this particular instance served to conceal the food item from conspecifics, then it may be classified as deception; and so on.

Behavior descriptions and their categorization were admittedly sometimes broad enough to could satisfy more than one of the five broad categories. For instance, it was not uncommon to find descriptions of social tool use, innovative tool use, innovative deception, among other combinations. To avoid conflation and thus overestimation of the strength of the positive manifold, such cases were not counted as duplicates, but rather only counted once if one aspect predominated (e.g., either tool use or innovation in the case of innovative tool use), or deleted altogether from the database if no clarity existed (Reader et al., 2011).

Secondary data collection was implemented by multiple researchers. After having established search, consideration, and collation rules, they worked independently for the first 10% of records examined in order to estimate inter-rater reliability. Martin and Bateson's (1986) index of concordance was used to this end. High agreement was attained, ranging from .83 to .95 depending on which of the five cognitive abilities was considered (Reader, 2003; Reader & Laland, 2001).

Though the present study employs Reader and coleagues' (2011) database, see the Discussion section for a proposal on future, long-term avenues of novel efforts for secondary data collection and therefore future refinements in what is possible to be estimated from the database.

Statistical Estimation of G

In previous literatures using this dataset, a *G* factor was estimated in an exploratory fashion using principal components analysis, principal axis factoring (Reader et al., 2011), and subsequently using unit weighted factor scoring (Fernandes et al., 2014). These approaches led to highly convergent factors. Even so, as factor loadings produced with principal components analysis and principal axis factoring in small samples are less reliable than those computed with unit-weighted factoring because of large standard errors (Figueredo et al., 1995; Gorsuch, 1983), the *G* factor used in the present analysis was derived from the unit weighted factor estimated by Fernandes and colleagues (2014). It explained 62% of the variance among the five cognitive capacities (see Chapter 2 for further details on the factor estimation and for factor loadings, which replicated the results of Fernandes et al., 2014).

2.1.2 Brain size

Data on 3813 specimens corresponding to 176 non-human primate species are available for overall brain size (Isler et al., 2008). Measurement is highly reliable, not needing correction formulae, given very high inter-researcher reliability in the estimations made (Isler et al., 2008). Moreover, for the overwhelming majority of data points, the original collecting locality and other information for the specimens are known, thereby avoiding misclassification with respect to sister species. Data from sources other than Isler and colleagues' own measurements were added from the literature by the original authors for species with insufficient data.

2.1.3. Neocortical, cerebellar, and hippocampal size

Stephan and colleagues amassed a database comprising the volumes of multiple neuroanatomical regions for 45 primate species. As is common practice for neuroanatomical analyses involving primate species (e.g., Deaner et al., 2007; Dunbar, 1992; Shultz & Dunbar, 2010), this database will be employed. However, it will be combined with recently published data made available by Navarrete and colleagues (2018) on more species and more specimens for many of the same species, totaling a 67-species database for the neocortex, hippocampus, and cerebellum.

2.1.4. Body mass

Data on body size, obtained from (Isler et al., 2008), permits residualizing NVMs against it to examine if it is relative or absolute NVMs that exhibit more evolutionary similarity to *G*. The updated data source compiled by Isler and colleagues will be used, as it not only includes original data for the same species as brain size, but also adds data points reported in previous

publications that focused on examining the validity of body mass measurement in primatology. Chief among these is Smith and Junger's (1997) effort to examine the shortcomings of previous sources commonly employed in comparative analyses, and to provide updated, more reliable estimates.

While it served the field immensely by motivating discussions about the neuroanatomical basis of intelligence for decades, the encephalization quotient will not be included in present analyses as a measure in analyses given (1) the now almost unanimous agreement upon its severe statistical limitations and biases, (2) its inferiority to absolute or residualized NVMs in its capacity to predict intelligence, and (3) its high variation in encephalization values assigned to species depending on which are included in analyses (for reviews, see Falk & Gibson, 2001; Peñaherrera-Aguirre et al., 2017).

For species with non-negligible sexual dimorphism, female measures will be used as is common practice (Isler et al., 2008), considering that the larger male size is generally mostly the result of intrasexual competition (Mitani et al., 1996; Plavcan & van Schaik, 1997) rather than an ecological adaptation for higher cognitive processing.

All variables will be log-transformed prior to analyses as is common practice in comparative studies, due to the observed high skewness inherent in cross-species data (Harvey, 1982). NVMs that are commonly residualized against body size in the cognitive literature will be included in both raw (i.e., absolute) and residual form in the analyses[2]. Residuals will be

[2] It should be noted that controlling for the so-called 'effects' of body size on the traits studied (on the basis that it may constrain/influence their evolution) is a common but controversial approach. It has been argued that controlling for size also removes adaptive

computed with ordinary least square regressions (OLS). The proportion of the neocortex to the rest of the brain (i.e., neocortex ratio), a common index in the comparative literature, will also be included in addition to raw variables and residuals. However, following Shultz and Dunbar's (2010) inclusion of the raw hippocampal volume only, and given the lack of other comparative studies that focused on residual hippocampal data specifically in general intelligence analyses, here the hippocampus volume will not be residualized against body size.

For all analyses, a phylogenetic tree will be obtained from *10ktrees.fas.harvard.edu* (Arnold et al., 2010). Phylogenetic trees represent the pattern of relatedness among species, with speciation events as nodes and daughter lineages that result from speciation as branches emanating from a node (for illustrative examples, see Figures 1.1 and 1.2; and Figure 2.2 from Chapter 2 for a real example from the primate order). Arnold et al. (2010) made available a consensus tree for the primate order, relying both on molecular data and fossil data available on the literature. Considering one goal of the present study involves estimating evolutionary rates of change on measures across time, the phylogenetic tree selected had branch lengths representing time elapsed since speciation (i.e., so-called ultrametric tree), with the horizontal axis of the tree reflecting time in millions of years. Furthermore, most phylogenetic comparative methods that will be used in the present study, including estimation of trait conservatism, have been developed

---

variance (stemming from adaptations to maintain functional equivalence or from a common cause of variation in size and in the trait in question; Fleagle, 1985; Jeschke & Kokko, 2009; Roff, 2011; Smith, 1980), thus reducing the power of the traits as predictors of intelligence when applied to NVMs (Deaner et al., 2007). Interpretations of results of residualized measures are done with caution, considering this caveat.

for this type of tree topology (Garamszegi, 2014). Data for traits of interest, to be analyzed using the phylogenetic tree through the methods described below, can be entered for extant species (i.e., at the tips of the tree). Rather than being independent data points, the data for the species have a pattern of interdependence determined by the tree topology, and this permits estimating (a) how conserved the trait in question is, (b) the rate of change for the trait across time (i.e., branch lengths for the tree), (c) the fit of several selection models to explain the extant species variation in the trait; as fully detailed in the section below.

2.2. Analyses

2.2.1. Phylogenetic signal

Phylogenetic signal (PS), frequently termed conservatism or inertia, refers to the degree of trait preservation across evolutionary time in a phylogeny. It reflects how much sister lineages are more similar to each other than to more distantly-related lineages on a given trait being analyzed. If a trait has been largely conserved and has not been too labile across evolutionary time, closely-related lineages will be highly similar to each other, more so than to distantly related species. Traits with null (i.e., 0) phylogenetic signal thus exhibit as much difference among sister lineages as they do among lineages that were separated further back in time, indicating that even in short evolutionary spans separating sister lineages, evolution has led to considerable divergence and the trait is thus highly labile. This suggests that a strong selection force has operated on at least some of the lineages, making them differ markedly from sister lineages. On the other hand, traits with a PS value of 1.0 exhibit high conservatism, meaning that sister lineages have not differentiated importantly on the trait. Note that high conservatism does not imply that the trait has absolutely *stagnated* and exhibited a complete lack of change

throughout all the phylogeny – this would be an unrealistic scenario as, across millions of years separating lineages, some changes are bound to occur. Instead, a trait is considered conserved (with PS of 1.0) if its changes across time are the result of a perfectly random walk (also termed Brownian motion; Felsenstein, 1985) where its values on the various lineages in the phylogeny go back and forth in no consistent direction. As in Brownian motion there is no consistent, directional selection on the trait but rather changes in random and varying directions, distantly related species will on average differ more markedly on the given trait than sister species that have just speciated, simply because distantly related species have had more time for random variations produced by the Brownian motion. As such, high PS values indicate that most species have not been under any important and consistent selection regime, although random changes still make lineages slowly diversify across millions of years.

Figure 1.1 provides an illustrative example of scenarios that would lead to low and to high PS estimates in body weight, an immediately relatable and easily understandable variable. While in Figure 1.1A values for sister species are more similar than when they are compared to more distant species (exhibiting PS close to 1.0), in Figure 1.1B values are just as dissimilar when comparing closely-related species as when comparing distantly-related species (exhibiting PS close to 0). Note that, in Figure 1.1A, sister species do not need to have identical values for PS to be considered high and conform to a Brownian motion expectation. Although the phylogenetic tree used in analyses in the present study are much larger, the rationale applies equally: Low values on PS indicate that, on average, sister lineages are just as dissimilar as distant lineages.

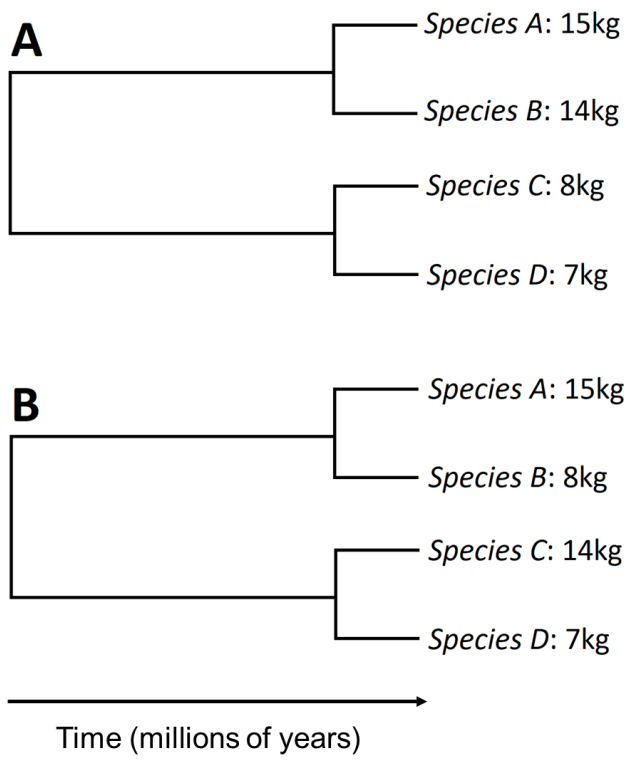

Figure 1.1. Illustrative examples of high and low phylogenetic signal. (A) High phylogenetic signal scenario for body mass for four hypothetical primate species: Closely-related lineages are more similar than distantly-related ones. (B) Low phylogenetic signal scenario for the same trait: Closely-related species differ markedly in body size, approximately just as much as they differ on average from distantly-related species.

Pagel's λ and Blomberg et al's *K* (Kamilar & Copper, 2013; Nunn, 2011) will be used to estimate PS for all variables in the present study. Both indices of phylogenetic signal will be used considering that there is no consensus in the literature as to which index better reflects the true pattern of conservatism (Münkemüller et al., 2012). *K* as a statistic tends to be an underestimate of true PS at low to intermediate levels, and λ tends to be an overestimate at intermediate to high levels, in simulation studies. As such, *K* and λ have produced considerably

divergent results in simulations and in empirical data (Fernandes, 2014; Münkemüller et al., 2012), sharing as little as 34% of the variance in estimated PS at times. A more accurate estimation of PS can be obtained using both metrics.

Moreover, *K* contains a unique feature: It permits the assessment of whether a trait is more conserved than expected under Brownian motion, whereas λ estimates range from absolutely no conservatism (0) to Brownian motion (1). A value higher than 1 therefore can be estimated for *K*, without an upper bound. Values higher than 1 indicate that close relatives are even more similar to each other than would be expected if a Brownian motion existed, suggesting that trait stagnation has taken place. Recall that Brownian motion is frequently a parsimonious model for most traits that are not under strong selection because it allows for random changes to slowly crop up in traits across time (Felsenstein, 1988); $K > 1$ on the other hand indicates that even random changes have been reduced or halted across evolutionary time, suggesting not only a lack of directional selection for changes but also strong constraint on any changes altogether. The higher than 1 the value for *K*, the stronger the stasis across time. This appears to be the case, for instance, of absolute brain size in primate phylogeny (Kamilar & Copper, 2013).

2.2.2. Evolutionary rates

Trait changes across time do not necessarily occur at the same rate for different variables. Traits under directional selection, for example, may exhibit rapid changes, making lineages differentiate from each other rapidly. Taking the average amount of difference on a given trait among lineages, and weighting it by the time span of the phylogenetic tree yields an estimate of the average evolutionary rate for the trait; that is, the amount of change per million years. Figure

1.2 illustrates how evolutionary rates may be inferred from tip data (i.e., extant species). Note that the estimates of evolutionary rates indicate the *average* rate observed for the phylogenetic tree as a whole, whereas estimation of the precise evolutionary rate at any particular point in time in the phylogeny would require data on ancestor lineages (i.e., data on specimens that lived in the past, between the root and the tip of the phylogeny, rather than solely using tip species data).

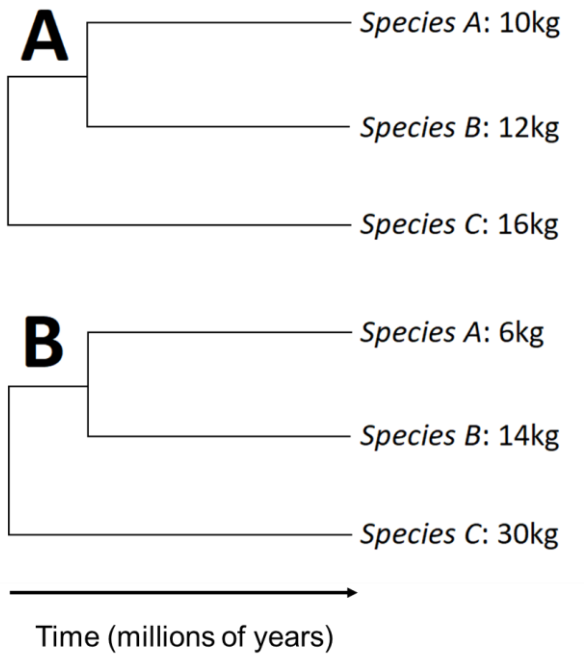

Figure 1.2. Illustrative example of the outcomes of two different evolutionary scenarios of evolutionary rates, upon the body mass of three species. (A) Low average evolutionary rate, reflecting small variance among species values in body mass. (B) Faster average evolutionary rate, reflecting larger variance among species values in body mass.

Potentially even under Brownian motion where trait changes occur in random and frequently change directions, trait changes may occur at fast rates, provided that the trait has little to no impact on fitness (even when its values vary widely). This has been demonstrated in simulations (Revell et al., 2008), but is an unlikely scenario with real data, as it is unlikely that a

trait (especially related to cognition or neuroanatomy) would permit wide and fast variations in random directions without impacts on fitness. As such, it is expected that, if fast evolutionary rates are identified for the traits in the present study, they will be on variables that exhibit a small phylogenetic signal (rather than the random walk of Brownian motion) and which exhibit signs of directional selection towards an optimum (see Section 2.2.3 for a description of how this will be assessed).

Evolutionary rates will be estimated with the Geiger package for R. Geiger estimates the average evolutionary rate for a given trait in the phylogeny (in this case, the phylogenetic tree for the primate order) by calculating the difference in scores among sister lineages and weighting that difference by the length of their branches (i.e., the time elapsed since they speciated), in millions of years. Then, taking the results for all pairs of lineages, the aggregate weighted difference is produced for that trait. This reflects how much, on average, primate lineages have changed per million years for the trait in question. NVMs and *G* can thus be compared in terms of their evolutionary rates by repeating this procedure for each measure described in the Measures section. Standardized scores (*Z*-scores, with evolutionary rates reflecting average number of standard deviations of change in each measure per million years) will be used as standardized scores are ideal for comparing traits that have different metrics (Hunter & Hamilton, 2002). Clearly, *G* and volumetric measures are not measured in the same metric, thus standardizing them permits a meaningful comparison. Note that, as previously described, all variables will be log-transformed for all analyses, a practice that also helps avoid overestimation of evolutionary rates for the traits with high average values (Adams, 2013; Gingerich, 2009; O'Meara et al., 2006). To protect against the possibility that the evolutionary rate reported for any variable is inflated by the standardization process, the coefficient of variance of all variables

was checked for comparability. No inflation was detected. The rationale, procedure and results for this check are described in Appendix B.

2.2.3. Selection regimes

As reviewed above, Brownian motion is a common approximation of the pattern of evolution behind traits. Recall that Brownian motion refers to the null model of evolution of a trait simply reflecting the passage of time, with no particular direction of trait change (i.e., consistent increases or decreases) and no alteration in the evolutionary speed (i.e., acceleration or deceleration), but rather reflecting a random walk under a stable rate (Nunn, 2011). Many traits in animal datasets and specifically in the primate order do not conform to Brownian motion evolution (Blomberg et al., 2003; Kamilar & Cooper, 2013). Testing for a phylogenetic signal shows if traits are not as conserved as would be expected under a Brownian motion model, but does not clarify *why* one trait exhibits high lability and another exhibits high conservatism in the phylogeny of interest; that is, simply testing for the phylogenetic signal of a trait does not tell us why they have diverged or failed to diverge from the Brownian motion expectation. To examine what selection regime led to divergence from the Brownian motion null expectation for any given variable, several models have been developed that each add a parameter to the null model testing the fit of Brownian motion to data (for a review, see Peñaherrera-Aguirre & Fernandes, 2018; Hernández et al., 2013). These parameters each test an aspect of evolution; for instance, one tests whether the trait exhibits a sign of having accelerated in evolutionary rate (thus becoming more of a target of selection), another tests whether the data for the variable conform to a pattern of niche filling whereby rapid change in the root of the phylogeny is followed by progressively more stagnation in evolutionary changes. While no model single-handedly

provides a detailed, comprehensive explanation of evolutionary changes in the trait in question neither do they provide nuances about microevolutionary specificities for any single lineage in the phylogeny, the models each test a possible scenario of the selection regime governing the general pattern of evolution for that trait. As such, they each signal about an aspect in the overall trend for the trait being analyzed, and help us understand why that trait exhibited low or high phylogenetic signal or low or high evolutionary rates. As each alternative model adds exactly one parameter relative to the Brownian motion model, the fit and significance of each can be tested using a likelihood ratio test.

Each trait will be examined under three evolutionary models: acceleration, early burst, and Ornstein-Uhlenbeck (OU), as detailed below (cf. Peñaherrera-Aguirre & Fernandes, 2018; Hernández et al., 2013).

I) Acceleration: Changes in a trait of interest may have become more essential to evolutionary success (i.e., survival and reproduction), having thus become more intensely the target of selection for modifications. This model tests whether the evolutionary rate for changes in a given trait has accelerated (Pagel, 1999). Relative to the Brownian motion model, it adds the parameter δ, which when larger than 1 suggests accelerated evolution towards the tree tips, and thus taxon-specific adaptations; values lower than 1 suggest deceleration over evolutionary time.

II) Early burst (EB): The possibility that lineages in a clade exhibit quick diversification in a trait followed by slowed change is a common hypothesis in ecology (Moen & Morlon, 2014) and may also lead to deviation from Brownian motion. This phenomenon is expected when new ecological niches open up and become saturated over time. The EB model (Harmon et al., 2010) permits testing this possibility by assessing slowed evolution. It differs from the previous model in that it estimates whether there has been an *exponential* decrease of the evolutionary rate of the

trait over time (Peñaherrera-Aguirre & Fernandes, 2018), by adding the parameter *a* relative to the Brownian motion model. When this parameter is estimated as zero it is equivalent to evolution under Brownian motion, whereas it is assumed that there is niche-filling (i.e., rapid change) followed by an exponential decrease of evolutionary rates if the value is less than 0.

III) Ornstein-Uhlenbeck (OU): An adaptive problem or a set of adaptive problems can sometimes exist for multiple taxa, which converge towards an adaptation. An optimal value (with directional selection towards it) may thus exist for a trait across taxa; for instance, pressure may have existed somewhat consistently towards a higher level of intelligence across many primate lineages. The OU model tests a parameter (α) against the Brownian motion null model, for the strength of a selection force; that is, a selection pressure towards a certain value for most taxa in the phylogeny (Hansen, 1997). Although often associated with stabilizing selection, the OU model actually simply examines whether the trait is being selected toward an optimum point, which can be achieved not only through stabilizing selection but also through directional selection (Ingram et al., 2012). The α values estimated by the OU model can range from 0 to infinity.

While a value of 1 for δ and a value of 0 for *a* and α mean perfect conformation of the data to Brownian motion for the variable being analyzed, no other standard values for cutoffs or comparisons exist, other than comparing the estimates against those found for other variables in the study, and testing their significance against the Brownian motion model. In other words, values range to negative and positive infinity, and their purpose is to be used to compare several traits, in addition to examining which significantly depart from the Brownian motion expectation. Likelihood ratio tests will be used to determine whether the fit of each model for

each measure was statistically different from Brownian motion. This will permit objectively determining if the models were a better fit than this null hypothesis.

3. Results

3.1 Phylogenetic Signal

The phylogenetic signal (PS) for *G*, cerebellum size, and neocotex size was low when estimating K, and low to medium when estimating λ. It was consistently higher (reflecting less evolutionary lability and thus stronger conservation) for other NVMs. These results are presented in detail in Table 1.1.

Table 1.1.

Phylogenetic signal estimates (Pagel's λ and Blomberg's *K*) reflecting the degree of conservatism of general intelligence (*G*) and of the neuroanatomical volume measures in the primate phylogeny

Measure	λ	*K*
G factor	.62*†	.13*
Residual brain size	.96*†	.34*†
Residual neocortex size	.44*†	.43†
Residual cerebellum size	.46*	.18*
Neocortex ratio	.88†	1.74†
Absolute brain size	.99†	3.31*†
Absolute neocortex size	.99†	1.46†
Absolute hippocampus size	.98†	1.00†

Absolute cerebellum size	.99†	1.68†
Absolute body mass	.99†	1.98†

Note: * denotes the parameter value is significantly ($p \leq .05$) different from the Brownian motion model; † denotes the parameter value is significantly higher than zero.

Only residual cerebellum size and *G* exhibited a *K* value not significantly higher than 0 (and in the case of λ, only residual cerebellum size). This indicates that only in these variables primate lineages have differentiated radically relative to ancestor taxa (and therefore to sister lineages), retaining only a small or negligible degree of conservatism, while other measures retain an important degree of conservatism in spite of millions of years of speciation. Recall that a PS value of 0 indicates that sister taxa are no more similar to each other than they are to distant lineages, indicating that selection completely erased or replaced the pattern of shared ancestry on the variable in question producing intense changes in most lineages (either towards converging values, or towards disparate values across lineages). Analyses of selection regimes presented subsequently in this Chapter will help elucidate what selection regime(s) is or are behind the high evolutionary lability of *G* and residual cerebellum.

Still, not only in *G* and residual cerebellum size, but rather in all residualized variables, λ was found to be significantly lower than the expectation of Brownian motion (PS = 1.0), suggesting that these variables exhibit at least some evolutionary lability. This was also the case for all residualized variables and the *G* factor when *K* was used to estimate PS, except for residualized neocortex size. Recall that, theoretically, variables perfectly conforming to Brownian motion exhibit a random walk of slow changes in varying directions, making sister lineages more similar to each other than they are to distant relatives. Such traits would be

considered conserved. In contrast, *G* and residualized NVMs exhibit a phylogenetic history with a significant sign of selection relative to the conservation assumption of the Brownian motion value of 1.0, although only *G* and residual cerebellar size exhibit PS not significantly different from 0.

A different picture was found for absolute (i.e., non-residualized) NVMs and body mass however, which not only exhibited the highest or close to highest possible λ values, but also surpassed 1.0 in *K* estimates in most cases, indicating that they are even more conserved than expected through Brownian motion. Recall that $K > 1$ suggests stasis or severe constraint in evolutionary change, leading sister taxa to be even more similar to each other than in cases where there is random slow change in varying directions; it appears that absolute NVMs conform to this pattern, especially brain size, showing a *K* value more than three times higher than would be found if it conformed to Brownian motion.

To summarize, the pattern of evolutionary lability estimated through λ and *K* suggests that variability with respect to absolute NVMs and body mass tend to be *selected against* or *constrained* in the primate phylogeny. Once controls for body mass are implemented for NVMs, it can be shown that there is some lability, but usually not as much as for *G* except for the cerebellum and, to a lesser degree, the neocortex.

3.2 Evolutionary Rates

In analyses of evolutionary rates, non-residualized *G* and residualized *G* exhibited higher rates relative to NVMs, which were almost all below .05, as displayed in Figure 1.3. The exception is cerebellum size, both with and without body size residualization: its evolutionary rates were approximately two thirds as fast as those for *G*. Although residual neocortex size and

residual brain size exhibited higher evolutionary rates than the remaining NVMs, their rate was only approximately half of that observed for the residual cerebellum. These results permit ranking of the examined NVMs into three main groups in terms of evolutionary rates: (1) residual and absolute cerebellar size were fastest, (2) residual neocortex and residual brain size were intermediate, and (3) other NVMs and body size were slowest, evolving up to an order of magnitude more slowly than *G*.

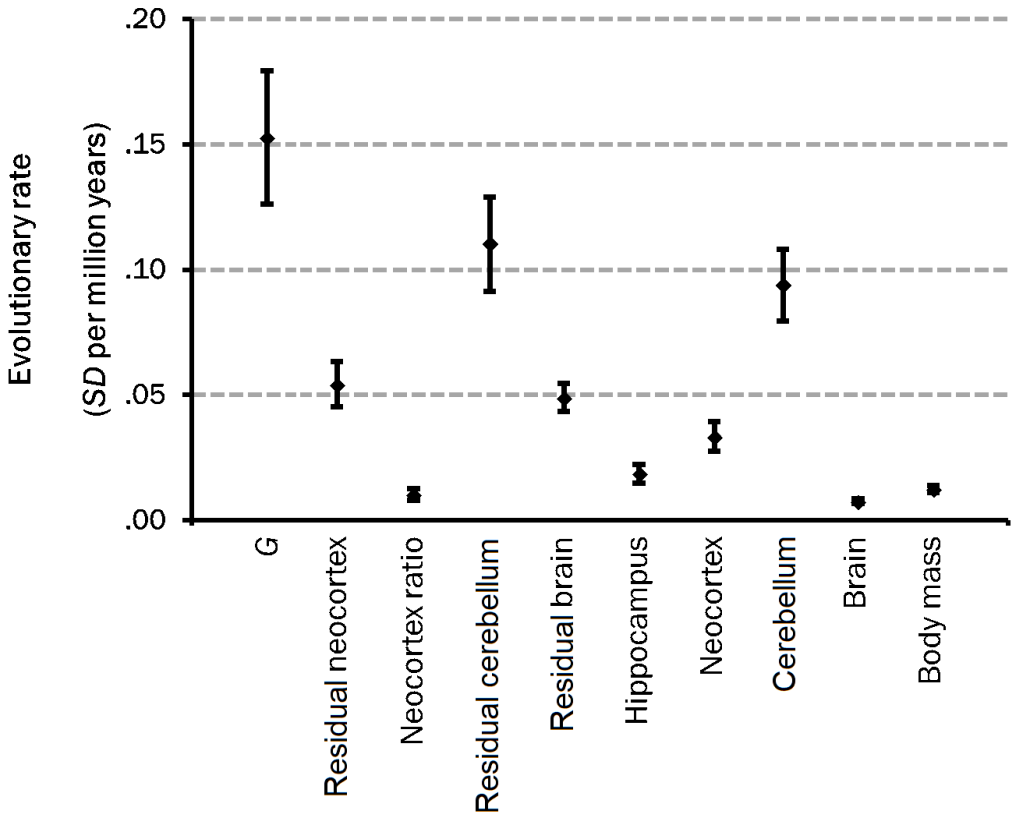

Figure 1.3. Estimated evolutionary rate in standard deviations per million years of *G*, of neuroanatomical volume indicators residualized and non-residualized against body mass, and of body mass, in a comparable metric after *Z*-score transformation. Bars represent standard errors of the mean.

3.3 Selection Regimes

A similar pattern of contrast between *G*, residual cerebellum and residual neocortex, and other measures, is suggested by the selection model comparisons, as presented in Table 1.2. The δ estimates were of high magnitude for *G*, around an order of magnitude higher than those estimated for residualized NVMs except for the cerebellum (presenting intermediate values), whereas non-residualized NVMs exhibited either no acceleration, or negative acceleration in the case of brain size (which is compounded by a negative, exponential acceleration identified with the *a* parameter of the early burst model). In accordance with this difference among measures, *G* exhibited a strong trend toward a selection optimum throughout the phylogeny, as can be observed with the α parameter of the OU model.

Table 1.2.

Parameter estimates for rate of acceleration, early-burst (EB), Ornstein-Uhlenbeck (OU), and phylogenetic signal (PS) models of evolution of species-level general intelligence (*G*) and of the neuroanatomical volume measures in the primate phylogeny

Measure	Acceleration (δ)	Early burst (a)	Ornstein-Uhlenbeck (α)
G factor	49.43*	.00	.34*
Residual brain	4.28*	.00	.03*
Residual neocortex	7.25*	.00	.08*
Residual cerebellum	19.74*	.00	.14*
Neocortex ratio	1.23	.00	.02
Absolute brain size	.19*	-.06*	.00
Absolute neocortex size	1.01	.00	.00

Absolute hippocampus size	1.43	.00	.00
Absolute cerebellum size	.59	-.04*	.00
Absolute body mass	.42	-.03*	.00

Note: The early-burst parameter *a* was constrained to have an upper ceiling of 0.0, as positive values represent the opposite of an evolutionary early burst, an expectation already tested in the Acceleration model.

Recall that the three selection models tested are not mutually exclusive, but rather they are tested against the assumption of Brownian motion in each trait. Identification of significant and very high estimates of δ and α for *G* and, to a lesser extent, for residualized NVMs (especially the cerebellum) indicate that in more recent primate history the rate of evolutionary change in these variables has increased compared to the rate estimated for the early history of primates, and that these accelerated changes have not occurred in random directions but rather mostly toward an optimum. Combined with results of previous analyses, it can be inferred that such accelerated changes toward an optimum were mostly *increases* in *G* and in the size NVMs relative to body size for most primate clades (Reader et al., 2011) though the present analyses by themselves are not designed to indicate direction, especially in the absence of data for ancestral species.

Additional analysis testing the relative weights of fit for the alternative models (Table A2.2, Appendix B) indicate that the acceleration and OU models are highly parsimonious for *G*. In contrast, the residualized NVMs assumption highly rejected the Brownian motion null hypothesis but the favored model was simply one of trait lability – indicating that they are not highly conserved traits but fail to exhibit a clear trend in a particular direction or in acceleration

or deceleration across evolutionary time. Again, the exception was residualized cerebellar size, for which the best fitting models were the acceleration and OU model just as in the case for *G*, although parameter estimates were not as high for the cerebellum as they were for *G* (see Table 1.2). Non-residualized NVMs, on the other hand, exhibited Brownian motion as the most parsimonious model, or in some cases they conformed to EB, with negative acceleration.

4. Discussion

This study aimed to compare the evolutionary patterns (i.e., phylogenetic conservatism, evolutionary rates) and processes (i.e., strength of selection regimes and changes therein across the phylogeny) of general intelligence (*G*) and neuroamatomical volume measures (NVMs) in the primate order. While numerous previous studies have assessed the correlation strength of *G* with NVMs, it has become clear in the phylogenetic comparative methods literature more broadly, that correlations do not necessarily imply shared evolutionary processes or evolutionary causation, with correlated traits not uncommonly exhibiting disconcerted patterns and processes of evolution (e.g., Fitzpatrick et al., 2012; Gonzalez-Voyer et al., 2009). While largely exploratory in nature, as a first examination of the strength of evolutionary convergence between *G* and the most commonly used NVMs, it was hypothesized that at least some differences would be identified, as (1) the case for other neural factors behind intelligence has been convincingly made multiple times, which would require that volumetric measures are not perfectly convergent with *G*, (2) the size of the brain or its components is frequently negatively associated with neuronal density, and (3) profuse debates over which NVM is the best proxy for *G* have led to the identification of several limitations and generally in moderate effect sizes.

It is clear that NVMs are employed as proxies for intelligence because of how easy it is to measure them, compared to histological indices, and because of the predictive power that size measures have upon some other neurological indicators. For instance, the encephalization quotient was of interest because of its hypothesized relation to the concept of "extra neurons" (above the number of neurons necessary to operate a body of the size of the species in question; Jerison, 1973). However, the multiple analyses conducted in the present study led to largely non-converging results when comparing *G* and most NVMs (with exceptions discussed below), suggesting a low similarity in their evolutionary patterns and processes. Overall, *G* appears to have been more evolutionarily labile, with faster and accelerating evolution that, on average, shifted the trait towards an optimal value rather than evolving at or close to a random walk. While in terms of evolutionary lability (measured through λ and K), two NVMs were comparable to *G* (residual cerebellum and neocortex volumes), multiple lines of evidence indicated that the evolutionary histories of NVMs are not highly comparable to that of *G*:

1) Evolutionary rates were found to be fastest for *G*, slow for absolute NVMs and slowest for body-size-corrected NVMs, suggesting the evolution of brain size and its components is at least partly tied to the evolution of body size in primates, and not remarkably fast. The evolutionary lability of *G* appears tied to a high evolutionary rate, a finding previously identified by Fernandes and colleagues (2014) for the specific cognitive abilities comprising *G* as well.
2) Deceleration of evolutionary rate was identified for brain size and body mass, and in contrast strong positive acceleration was found in the case of *G*. Residualized NVMs exhibited comparatively small or null acceleration. Again, this suggests that, for *G*, evolution at the tips of the tree has been strong, as opposed to the case of NVMs,

especially when they still retain variance associated with body size. This body-size related deceleration of evolutionary rate in primates confirms previous findings by Cooper and Purvis (2010).

3) Weaker or null selection trends were found toward an optimum for controlled and uncontrolled NVMs, whereas a stronger trend toward an optimum value was found in the case of *G* (as measured by the Ornstein-Uhlenbeck model parameter α). As argued by Revell and colleagues (2008), selection toward an optimum can lead to high evolutionary lability in the phylogeny, which was empirically confirmed in the case of *G* in primates.

Cerebellar and, to a lesser degree, neocortical volumes, when residualized against body mass presented the most similar model fit results compared to *G*, in addition to similar phylogenetic signal estimates, even though there were still striking differences in most parameters – in multiple cases parameter estimates being more than twice as large for those found for cerebellar and neocortical volume measures. It is counter-intuitive that non-residualized brain size exhibited the least comparable parameter estimates and model fit results relative to *G*, while an important correlation exists between these two variables (Deaner et al., 2007), a point which further compounds the low generalizability in the interpretation of correlation coefficients, in that they may poorly reflect underlying evolutionary processes. It must be noted that, while brain size and *G* may be moderately correlated, over evolutionary time the proportional changes in *G* size appear much larger than the proportional changes in brain size. As such, while changes appear to occur in somewhat converging directions across evolutionary time for these two traits, the *amount* of modification observed for *G* appears more than an order of magnitude higher, with brain size being, in contrast, much more conserved.

It is possible that correlations of NVMs with *G* may reflect *constraints* imposed by the former upon the development of cognitive abilities. Their generally modest correlation magnitude may indicate that a given size of an NVM accommodates a wide range of values in cognitive abilities, but beyond that window increases in volume are necessary for further increases in cognition. This possibility is akin to the proposed relationship between body mass and brain size itself (e.g., Gonzalez-Voyer et al., 2009). As such, it would not be invalid to use NVMs as a proxy for intelligence in the absence of cognitive data given their considerable phenotypic correlation, so long as there is awareness of the increasingly clear limitations of this approach when dealing with their evolutionary interpretations.

4.1. Moderate Similarity Between G and Residual Cerebellar, and to a Lesser Extent, Neocortical Volume

It is striking that, of all NVMs and their operationalizations examined, cerebellar size residualized against body size displayed the most similar results to those for *G*. Phylogenetic signal for the cerebellum was significantly different from 1 (i.e., from the assumption of conservatism) just as in the case of *G*. While their evolutionary rates and fit to selection regime models were not identical, residual cerebellar volume appears to have evolved faster than other NVMs, and exhibited considerable acceleration and a selection trend towards an optimum. The fact that rates, acceleration, and trend towards an optimum were all somewhat lower than *G* but not radically divergent suggests that, while this neuroanatomical structure is not a sufficient substrate for *G*, it may serve as an important substrate. This seems especially plausible considering recent evidence that cerebellum size has changed in lockstep with overall cognitive ability (Smaers et al., 2018), and also specifically with technical or physical aspects of

intelligence (Barton, 2012). Cerebellar size also exhibited rapid expansion in great apes (taxa that have high *G*; Deander et al., 2006; Reader et al., 2011), more so than the neocortex (Barton & Venditti, 2014) while exhibiting less reduction in neuron density (Barton, 2012). It is possible that technical intelligence, requiring cerebellar specialization (given its role in sensory-motor control and in learning complex movement sequences), was central to the evolution of intelligence. This is in line with findings by Fernandes and colleagues (2014) and novel findings presented in Chapter 2 that tool use exhibits the strongest factor loadings in primate *G* and is the ability most correlated to others in species of high *G*, among the five abilities examined. Tool making involves carefully organized, well-coordinated motor-sensory sequences and having a template of these actions in mind; the cerebellum (in connection with cerebral sensory-motor areas) is a strong candidate substrate for this capacity. As such, it has been argued that, under certain ecological circumstances present in the evolutionary history of some primate taxa, the evolution of higher connectivity among regions involving executive, perceptual, and motor regions was necessary for complex cognitive abilities such as innovation (Navarrete & Laland, 2015). It is also possible that the cerebellum functions as an *augmenter* of the activity of other brain structures (Leiner et al., 1989; Snider, 1950); as such it would enhance cognitive skill when projecting to regions largely responsible for it. As reviewed subsequently in Chapter 2, the evolution of larger brains in primates has been accompanied by an expansion in white matter, with increased myelination compensating for the long range of connections; it is likely that this has permitted the cerebellum to not only maintain but arguably increase its influence on other, cortical areas, facilitating the motor coordination characteristic of intelligent problem-solving.

Surprisingly, relative to the number of studies examining or proposing a role of overall brain size or neocortical size in intelligence, the role of the cerebellum is extremely understudied

and understated. Further attention to cerebellar size in relation to cognition in primates is warranted.

It is also somewhat puzzling that neocortical volume residualized against body size is a rarely used operationalization of neocortical size in comparative studies, with researchers instead relying on the neocortex ratio (e.g., Dunbar, 1992; Shultz & Dunbar, 2010) or overall brain size-related measures (e.g., Deaner et al., 2007; Gibson et al., 2001). The results of the present study suggest that relative neocortical volume is, second to residual cerebellar size, the most similar to *G* in terms of low phylogenetic conservatism, intermediate evolutionary rate and rate acceleration, and some sign of evolutionary changes toward an optimum size. While the role of the neocortex is cognition is well-discussed in the comparative and human cognition literatures as reviewed at the outset of this study, further attention should be given to how it is operationalized in comparative studies, considering that the ratio approach exhibits high divergence from *G* in their evolutionary patterns.

4.2. Putative Alternatives to Volumetric Measures

Questioning the application of size-related neuroanatomical measures to understanding intelligence is not in itself a novel endeavor. The idea that the reorganization of systems internal to the brain reflected evolutionary changes in cognitive abilities better than the size of the brain of subcomponents is not uncommon or recent (Holloway, 1966a,b). More recent reviews of the literature point to a plethora of studies that indicate how evolutionary reorganizations of the cortex are common (Preuss, 2001).

How might *G* have been selected for across primates, net of brain volume? Even metrics that had results most comparable to those of *G*, such as residual cerebellar size, were still

noticeably different in evolutionary rates and the fit of selection regime models, not appearing as sufficient substrates for the remarkable evolution of primate intelligence. There are several possible scenarios proposed and explored in animals in general and specifically in primates as well. These alternatives are more than likely complementary to each other as substrates for intelligence, and complementary to the low to moderate role of NVMs identified in the present study, rather than full substitutes. The alternatives outlined below appear as highly promising future avenues for further research, although limited amounts of data are available for analyses on primates at the current moment. A small number of data points exist for non-volumetric potential correlates of *G*, outright preventing their immediate use in analyses as phylogenetic comparative methods require a minimum of approximately 15-20 species for sufficiently reliable estimates.

1) *Gyrencephaly*, which is characterized by an increase in the degree of convolutedness of a brain, could have increased the surface area available to accommodate more complex neuroanatomical structures and attendant cognitive systems without having to proportionately increase volume (although some increases are necessary and observed, and thus a partial positive correlation exists; Gibson et al., 2001). Gyrification may also have evolved to operate in combination with other features. For instance, folding may reduce connection length among cortical areas (Hofman, 2001).

2) An increase in the degree of myelination, which facilitates increased information processing speed, may be another factor behind intelligence. As with humans (e.g. Jensen, 2006), differences in glial density and myelination (both involved in processing efficiency) should be comprehensively examined across samples of primate taxa. Preliminary evidence comparing species qualitatively suggests an important role of myelination in cognitive ability (Dicke &

Roth, 2016), as well as a role of progressive myelination in the maturation of cognition within species across taxa (Gibson, 1991).

3) Processing power, so far a largely neglected area in intelligence studies, is also a potential candidate. While it would be extremely unlikely that any single localized area that has evolved in a largely independent fashion would be responsible for general intelligence (given the necessity of this domain-general process to recruit from a multitude of cognitive resources), several interconnected localized areas, working as a circuit, may be candidates for explaining the evolutionary trajectories of general intelligence. While the overall correlation between brain size and neuron density is negative (small to moderate), there are excepted areas that may be of interest to studies of intelligence evolution, such as area 10, where there is relatively more variation off the allometric line (Semendeferi et al., 2001). Moreover, candidate areas identified from within-species analyses (e.g., Colom, Jung, & Haier, 2006; Duncan et al., 2000; Haier et al., 2004, 2005) should be of potential to comparative analyses.

4) Cortical microanatomy should also be examined. New interneural projections for increased connectivity in complex networks that accommodate intelligence would necessitate enlargement of existing pyramidal cells (or the generation of new pyramidal cells) in the areas from which projections are made, in order to support new axon collaterals (Preuss, 2001).

Dicke and Roth (2016; Roth & Dicke, 2005) make a compelling case that the best fit between brain traits and intelligence in animals, at the cross-species level, involves a combination of several factors that determine general information processing capacity, such as the total number of cortical neurons, neuron packing density, interneuronal distance, and axonal conduction velocity, in addition to other factors such as pulse width, gyrification, and differential allocation of connection to nearby versus distant areas (cf., Hofman, 2001). As such, there would

be no single measure that serves as a substrate for intelligence and therefore represents it, but rather a collection of integrated features such as those listed above, preventing exponentially costly increases in any given single factor that permits processing capacity.

Caution rather than excessive assertiveness about the explanatory power of these alternatives is necessary before further empirical, comparative evidence, as brain and body size tend to overall correlate positively with several of these alternative measures, such as ratio of connections to neurons, numbers of gyri and fissures, size of several specific brain regions, and cerebellum (Gibson et al., 2001). As such, the common practice of employing controls and examining residuals when dealing with NVMs may be extended to these alternative measures. Moreover, the plausibility of any of these alternatives and possible future confirmation of their roles as substrates for intelligence does not negate the partial relation that NVMs have with *G*: The most essential point to be made is that NVMs are not a *sufficient* explanation of *G* as only some of the NVMs examined in the present study moderately replicated the evolutionary patterns and processes observed for *G*.

4.3. Limitations and Future Directions

Although used in multiple research programs due to its demonstrably high correlations with experimental data and qualitative rankings based on expert analysis, the ethological count approach to estimating *G* is not without its limitations. It rests upon the observed frequency of only five indicators of high cognitive ability and relies on controls for research effort as different species have received different amounts of scientific attention by research groups – some species exhibit an extreme paucity of data. Ideally, a larger number of indicators would be collated, and

a more systematic effort for uniform attention across taxa would exist, however such a concerted effort unfortunately does not yet exist.

Nevertheless, it is extremely unlikely that the striking results observed in the present analyses were simply a function of measurement error. This is because random error is, contrary to systematic error, by definition likely to exist in all directions rather than consistently driving results toward a particular, specific trend. Were low PS (i.e., high disagreement among sister clades, possibly reflecting error in measurement) identified along with no clear evolutionary process behind it (i.e., null parameter estimates for the OU, acceleration, and EB models), a hypothesis about random error in measurement being responsible for results would be reasonable. However, in the present study, along with acceleration of evolutionary rates across time, a strong and consistent selection regime *toward an optimum* (i.e., with a direction) has been identified, which militates against the possibility of low reliability in trait estimation. Moreover, even though standard error of the mean estimates for *G* (e.g., in evolutionary rates) are larger than for NVMs, the average difference is so large (at times surpassing an order of magnitude) that any overlap in estimation is extremely unlikely.

The point must be made that in comparative studies of NVM variation across species have been primarily limited to volumetric data published in few studies and are often based on measurements of very few specimens of each species (Stephan et al., 1981; Frahm, Stephan, & Stephan, 1982) except for the case of overall brain size. The amount of studies reanalyzing these data is surprisingly high (for a review, see Herculano-Houzel & Lent, 2005). As in the case of expansions and revisions of body mass data (e.g. Smith & Jungers, 1997), obtaining further data on these regions is imperative, and the inclusion of novel data by Navarrete and colleagues (2018) in the present study helps reduce bias. Furthermore, considering the lack of convergence

in the results presented in this study, it is recommended that neurological measures other than those of a volumetric nature be given future attention. Several of the alternatives outlined above are already known to be largely independent of NVMs (Herculano-Houzel & Lent, 2005) and may thus serve as good sources of *complementary* information for understanding the evolution of intelligence.

Preliminary analyses of clade-specific evolutionary patterns and processes *within* the primate order (e.g., comparing suborders, superfamilies, etc.) suggest that a strong agreement between *G* and NVMs is not found either (Fernandes, Peñaherrera-Aguirre, Woodley of Menie, & Figueredo, in preparation) and that moderate comparability of results for *G* is found only with residual cerebellar volume, much like the general results detailed here, in addition to comparability with the hippocampus. However, interesting trends arise that complement and inform previous research: For *G*, the largest selection regime difference appears to be that existing between apes and other primates, whereas (with the exception of residual cerebellar volume and absolute hippocampal volume) in most residualized and non-residualized NVMs, the largest selection regime difference appears to be between simians and prosimians, or between catarrhines (apes and Old World monkeys) and other primates. This compounds arguments by Gibson and colleagues (2001) to the effect that measures such as the neocortex ratio are not ideal for distinguishing apes from monkeys in terms of cognition, while lending partial support to some level of comparability between the hippocampus and general intelligence (Shultz & Dunbar, 2010) and to the already discussed moderate comparability between cerebellar volume evolution and the evolution of *G*. This should nevertheless not be exaggerated in view of the moderate discrepancies found even between *G* and the most comparable NVMs, cerebellum

included. Further analyses that delve into clade-specific comparabilities and discrepancies within the primate order should help illuminate the neural bases of general intelligence.

4.3.1. Future database refinement: Collation of data on circumstances for behaviors, and categorization of subtypes of the five cognitive capacities, for the development of a taxonomy of behavioral problem-solving

For decades, a differentiation has been made between how traditional psychometric approaches to personality and cognition see these traits, and how a behavior assessment approach sees them. While traditional approaches view personality and cognition as traits the individual has, the behavior assessment literature focuses rather on what the individual does (Hartman et al., 1979; Mischel, 1968). Instead of picking one of these schools, the studies that originated the database used here were interested in what organisms of a species have the *capacity to do* (Byrne, 2003; Reader & Laland, 2002; Reader & MacDonald, 2003; Reader et al., 2011). These databases hold counts of how many times each species was observed to execute a broad kind of behavior. Future steps are needed to improve the level of detail of what species have the capacity to do. Psychometric and behavior assessment approaches are successful precisely because they have not remained this simple. Specifically, the database would benefit from detail on:

1) How the five broad categories of problem-solving can be subcategorized into more specific forms of behavior;

2) How complex the behavior is observed to be in each species (note that different species can execute it to different levels of complexity, or with different efficiencies);

3) Whether the species exhibits the behavior in all or in only few environments, and, in the case of the latter, what kinds of environments and conditions;

4) The frequency of the behavior in each species;

5) Whether few or most individuals exhibit the behavior in each species;

6) The ease with which a complex type of behavioral problem-solving is learned in each species;

7) The capacity or lack thereof for some behaviors (e.g., pointing to convey information, fashioning complex tools with fine motor dexterity) in each species.

The current form in which Reader and colleagues' (2011) database exists does not permit detailing what specific kinds of behavior were examples of each of the five broad categories of behavior problem-solving, and therefore does not permit developing a taxonomy of behavior (where broad categories contain subcategories, and so on down a hierarchy, until specific behaviors are detailed). Note that, as already explained in the Method, to avoid subjective bias, each behavior description found in the literature by Reader and colleagues were included in one of the five broad categories based on each original author's descriptions and classification, rather than based on any set of objective rules established by the researchers collating data onto the database. An important step towards the development of a taxonomy of problem-solving behavior is to establish overarching rules to permit a systematic categorization conducted by secondary data collectors (i.e., collators of the database) themselves. Such rules need not be subjective in the figurative sense of being superficial, imprecise, or not agreed-upon; rather, they can emerge from organized focus groups containing experts in the field. It is likely that some records of behavior that exist in the literature will not be amenable to classification into the taxonomy as many lack detailed descriptions (rather simply containing the original authors' classification of what category of problem-solving the behavior was an example of). Even with

such losses in the numbers of usable published records of behavior, the initiative to develop a taxonomy with clear overarching rules for the classification of behavior should advantageously trade off, by reducing the number of usable studies somewhat but adding detail, precision, and therefore the potential to further understand the evolution of complex problem-solving.

Regarding aspect number 1 in the above enumeration of improvements, it must be noted that a few studies using the same approach as employed in the database used here have already attempted to identify subtypes of some of the five broad categories of problem-solving, thus leading to a taxonomy of behavior. Whiten and Byrne (1991), for example, compiled a comprehensive taxonomic organization of tactical deception in primates which is readily usable for analyses (see Appendix A for a characterization of the categories of tactical deception behavior). However, other endeavors were either (a) based on within-species analyses of a single species (see the taxonomy of extractive foraging behavior based on snub-nosed monkey behavior detailed in Appendix A); or (b) limited to other taxa (e.g., birds) only, and to innovation and tool use, so far not including extractive foraging, social learning, and deception. As many as 12 subcategories of innovation, and 5 of tool use, were extracted in a bottom up approach based on the ethological descriptions found in the avian literatures (Boire et al., 2002; Overington et al., 2009). To illustrate, subtypes of tool use included tool use for feeding and drinking, and four forms of borderline tool use: dropping prey on a hard substrate, battering on an anvil, baiting, and holding prey with a wedge or skewer; Boire et al., 2002). Developing such taxonomies permits a thorough analysis of extant data (e.g., the phylogenetic analyses conducted for the five broad categories of problem-solving in the present paper could be equally conducted on each subcategory). The behavior taxonomies also permits the identification of unique or shared neural

substrates among the categories, when data on subcategories are compared to neural data using phylogenetic correlative methods.

This does not mean, however, that a category is by nature always limited to the number of subcategories identified in a data collation; for example, tool use is not necessarily limited to these 5 categories in avian species. If a new form of tool use were described in the literature, such as tool use for self-grooming behavior, a thirteenth category would emerge from the data without impediment. Similarly, if a new type of innovation were described in the literature, a new type would be included in the taxonomy. It is especially important for hypothesis-driven ethological observations to test the existence of (possibly partly) independent forms of tool use, innovation, and of other kinds of complex problem solving behavior. Such a combination of bottom-up and top-down approaches to the outlining of a taxonomy of behavior would be most powerful in avoiding false positive and false negative (sub)categories.

Having measures that contain gradations of frequency, complexity, ease of learning, among others mentioned above, one can compare, on each of these aspects:

1) Between-species comparisons – thus using the normative approach, where central tendencies are identified for clades and analyses reveal how species differ from them. This permits analyzing the macroevolution of the traits in question.

2) Within-species characterization– thus using criterion-referenced testing, where each species is understood in isolation relative to cutoffs in the traits' scales. This permits assessing the degree of cognitive complexity of each species, without requiring macroevolutionary analyses, rather relying simply on comparing scores to arbitrary or established thresholds, much like the edumetric approach used in education research (Carver, 1974; Hartman et al., 1979).

Furthermore, for some of the five broad categories of problem solving, but not all, data on the identity of the individual performing the behavior patterns and the circumstances surrounding each recorded behavior are described in Reader and Laland (2001). However, this is an outdated and incomplete database of even the few categories included, compared to the one used in the present studies. Future effort to detail the circumstances of behaviors would enhance the capacity for the development of a taxonomy, as it would facilitate understanding the purpose a behavior served: Returning to the well-known example of potato washing, knowing the conditions in which a behavior occurred permits understanding if the food item was immersed in water as a novelty (functioning as an innovation), or as a means to conceal it from conspecifics (functioning as tactical deception), or if a tool was involved (functioning as tool use), among other possibilities. While the act itself of immersing the food item in water is the same in all these circumstances, the function it serves varies.

Future studies potentially relying on a refined database as per the recommendations forwarded here would avoid capitalization on chance if a *confirmatory* factor analytic approach were used with newly compiled data. Reader and colleagues (2011) and Fernandes and colleagues (2014) already employed exploratory techniques and revealed the main five categories of complex problem solving, with such publications serving as pilot both in terms of data compilation and identification of the main, broad factors. Though ethograms commonly lead to behavioral *counts* (especially if they are derived from a one-zero recording technique), they are amenable to confirmatory factor analysis (CFA) as well as they involve quantitative data (Figueredo et al., 1992). Although unusual in the ethological literature, this particular study employed CFA to examine if factors extracted from data in fact confirmed the taxonomic

organization of behavior employed in the ethogram. In the revisions proposed here for the database, data recorded may not involve only counts, but rather continuous accounts of the behaviors (i.e., alternatives to one-zero recording), permitting, with even more accuracy, factor structure estimation using CFA. For revisions of the database, it would be especially useful to only include the primary studies with published and high inter-observer reliability, avoiding anecdotal records and initial pilot studies that relied on unstructured observations rather than on ethograms. Combining this set of rules for the new secondary data collection with confirmatory factor analysis would provide the scientific rigor expected of this growing area of study (Burkart et al., 2017).

In sum, limited comparability is currently found in terms of evolutionary trajectories of *G* relative to NVMs. Nevertheless, it is also important to compare *G* and NVMs in another sense: whether they are similar in the degree to which their internal structure changed over evolutionary time. In other words, it is possible that the strength of the manifold among cognitive abilities in *G* changed across the primate phylogeny, and it is also possible that the strength of the manifold among sizes of brain regions similarly changed. Chapter 2 will examine this possibility in *G*, and Chapter 3 will examine it with data on brain regions. The comparability between changes in internal structure of *G* and those in the internal structure of the brain will then be discussed and integrated with findings presented in this Chapter.

Chapter 2. Examining the Strength of Relations among Cognitive Abilities across the Primate Phylogeny: Clades of High Levels of General Intelligence Also Exhibit a Stronger Manifold

1. Introduction

Previous studies have tested the hypothesis that cognitive abilities in primates are organized in an integrated way, as reviewed in the Introduction and in Chapter 1 (see also Burkart et al., 2017). Nevertheless, they have not examined or demonstrated *how* the *G* factor of general intelligence has evolved – has it involved increasing levels of integration? Or was a *G* factor with many correlated abilities already present in the root of the primate phylogenetic tree?

In the comparative literature, general intelligence is argued by some to allow an organism to excel in a wide variety of contexts requiring behavioral flexibility (Deaner et al., 2006). However, it is irrational to expect evolution to shape phenotypes that serve as adaptations to problems beyond the ecology of an organism's ancestors. As such, along with many theoreticians, De Waal (2016, p. 12) has argued that "ranking cognition on a single dimension is a pointless exercise. Cognitive evolution is marked by many peaks of specialization. The ecology of each species is key." Nevertheless, an important realization is that in their ecologies a considerable portion of vertebrate species face a diversity of challenges and, most importantly, *novel* challenges. In other words, their ecology is not conducive to specialization (Parker, 1978). As such, zoologists and ecologists have long attempted to understand what characteristics generalist species possess and how they are able to cope with multiple conditions, with general intelligence being an important candidate.

1.1. The Evolution of Correlations among Cognitive Capacities

Because ecological specialists are able to use specific resources efficiently, they are favored in stable environmental conditions, whereas unpredictable or complex environments tend to favor generalists, as they do not rely on few specific, but rather on several diverse resources. Although behavioral generalism is not paid attention to as frequently as dietary generalism in literatures examining the ecological generalism-specialsm continuum, ecological generalism has long been proposed to be linked to behavioral flexibility, diversification of behavioral output, lack of neophobia, and proactive exploration of resources and of ways to access them. Empirical demonstrations of this relationship are longstanding (Glickman & Sroges, 1966; Parker, 1978; Visalberghi & Mason, 1983).

Indirect evidence suggests that general intelligence may indeed be important for coping with novel challenges, serving as a generalist adaptation. Game-theoretic models suggest that the existence of novel challenges for resource acquisition selects for cognitive flexibility capacities that permit generalism and innovative foraging (Overington et al., 2008). In line with this, comparative studies using avian species indicate that the capacity for technical innovation (e.g., through novel searching and handling techniques during foraging) is a hallmark of generalism, predicting invasion success (e.g., through range expansion or introduction to new habitats; Lefebvre & Sol, 2008) and dietary and habitat breadth (Overington et al., 2011). Innovation rates and extractive foraging have also been shown to highly correlate with other technical abilities and social abilities in primate species, giving rise to a general factor of intelligence at the comparative level (Fernandes et al., 2014; Reader et al., 2011).

A large piece of the scientific puzzle regarding the putative association between intelligence and generalism is missing, however. If general intelligence is an adaptation to

ecologies that select for generalization rather than specialization, the internal structure of general intelligence itself should reflect this. Species that have undergone selection for higher general intelligence should exhibit stronger *integration*[3] among various cognitive abilities than other species; that is, stronger intercorrelations should be observed among measures of problem solving, when examining problems of a different nature. If, on the other hand, species that exhibit high scores on general intelligence (e.g., as measured by the five problem-solving tasks in Reader et al., 2011) show weaker correlations among the cognitive abilities that constitute the general factor than others, it could be understood that their intelligence is rather specialized. In other words, in this case their high average scores would have been driven by performance in one or a few domains, rather than being generalized.

While factor analyses of performance on different cognitive problems have already been conducted and show the existence of a general factor among primate species (as reviewed in the Introduction and Chapter 1), factor loadings reflect only the *average* association among variables in the sample being studied. Other techniques are needed to examine if the strength of associations varies among species (e.g., with higher factor loadings as a function of average general intelligence). Thus to understand if species of high general intelligence exhibit integration among cognitive abilities, one needs to combine factor analysis with techniques that

[3] The use of the term "integration" does not imply *how* cognitive abilities are associated, or that they are integrated in any clear neuropsychological way. It refers simply to stronger sets of correlations among abilities. How these stronger associations among abilities take place in the brain is also a matter of research.

permit datum-level estimation of factor loadings, such as the Continuous Parameter Estimation Model (CPEM; Gorsuch, 2005), as will be defended and detailed in the Method Section.

1.2. Objectives and Hypotheses

This study aims to examine whether the strength of factor loadings in a cross-species (i.e., comparative) general factor of intelligence in primates varies as a function of the general intelligence score itself. While CPEM has been used to examine if interrelations among indicators in a given common factor vary as a function of a third variable (or as a function of scores on the common factor itself) in humans, for intelligence and other psychological variables, this is the first application of this approach to cross-species analyses.

The main hypothesis forwarded in this study is that higher general intelligence scores would predict higher factor loadings on the general intelligence common factor. As general intelligence is theorized to be an adaptation for novel, complex problem-solving in generalist species, it is predicted that cognitive abilities should be integrated, rather than exhibiting low factor loadings (i.e., being largely independent and thus specialized).

An alternative hypothesis, namely that the factor loadings would be equally strong across taxa, would necessitate revisions to the ecological specialism and generalism literature. This alternative would mean that, irrespective of ecological conditions, species would exhibit the already-demonstrated moderate intercorrelations among cognitive abilities, having less than 50% of variance left for specialized abilities (as more than 50% were explained by the common, general intelligence factor; Fernandes et al., 2014; Reader et al., 2011).

Finally, the polar opposite alternative that factor loadings would be weaker in species of higher general intelligence than in others would reduce the plausibility of general intelligence as

a relevant phenotype for understanding the nature of problem solving. This alternative would mean that it is exactly in species that exhibit low general intelligence that performance in problem-solving tasks correlate highly among themselves. As such, whenever a lineage evolved higher capacity to solve problems in one domain, it would be unlikely that the ability could be generalized to solving other kinds of adaptive problems. In other words, the existence of a strong manifold among problem-solving performance measures, in this scenario, would simply reflect a generalized *inability* for novel and complex problem-solving on the part of species of low average ability. It is argued that this scenario could not serve as support for the existence of a general factor of intelligence or for its reflecting ecological generalism.

2. Method

2.1. Measures

For consistency with Chapter 1, and considering their already demonstrated validity and robustness, the same five ethological measures of performance on cognitive domains utilized in Chapter 1 (originally compiled by Reader et al., 2011) will be employed for the present analyses. The *G* factor extracted from these abilities by Fernandes et al. (2014) will be used, just as in Chapter 1. Contrary to Chapter 1, where the abilities were not analyzed individually but only the *G* factor was used, the present study will use both on the five measures and on *G*. For a description of the five abilities, see the Method section of Chapter 1; for their expanded conceptualization, a review of existing operationalizations used in the original ethological observations compiled by Reader and colleagues (2011) for the database used in the present work, and behavioral examples, see Appendix A.

2.2. Analyses

2.2.1. Estimating integration coefficients

The *G* factor will be entered as a predictor of each of the five vectors of factor loadings (each vector corresponding to the part-whole correlation of each cognitive ability with the common factor), each in a separate analysis, totaling five analyses. This will permit testing whether species higher on the *G* factor exhibit a stronger manifold among abilities. The vector of factor loadings for each of the five cognitive abilities will be implemented with the Continuous Parameter Estimation Model (CPEM; Gorsuch, 2005; see also Figueredo et al., 2013 and Woodley of Menie et al., 2015, for detailed mathematical explanations of CPEM).

CPEM permits the change in the covariance between two variables (such as a higher-order construct and one of its indicators; in other words, a factor loading) to be correlated to another variable. As such, rather than leading to a single factor loading estimate for the whole sample as in traditional factor analytic methods, CPEM allows one to estimate variations in factor loadings throughout the sample, and permits these variations to be predicted by another variable. For all cognitive abilities, continuous parameter estimates of factor loadings will be estimated by taking the cross-product of the standardized (*z*) scores of each species performance on the task with the common factor.

When using more traditional methods (such as sample polytomization) for identifying changes in the strength of the correlation coefficient or of factor loadings between groups of species, it is necessary to acquire data for at least 75–100 species, so as to stabilize the correlation coefficients for comparison. On the other hand, as a graded method, CPEM does not necessitate the polytomization of continuous distribution (i.e., a division of the dataset into subgroups).

Traditional factor analysis techniques permit estimating the *average* factor loading of the general factor upon each measure. CPEM will permit the estimation of the factor loadings of the *G* common factor upon cognitive capacities for *each* of the 69 primate species for each measure. These species-specific factor loadings will then be predicted by the general intelligence scores of species to test the research hypothesis. The main hypothesis will be supported if species with higher general intelligence exhibit stronger factor loadings; that is, stronger associations among the five cognitive abilities. This effect of *G* upon factor loadings is called "cognitive integration coefficient", a term used throughout this text, as the predictive power of the common *G* factor upon species-specific factor loadings reflects the degree to which more intelligent species exhibit more strongly associated abilities. For clarity, Figure 2.1 illustrates how the latent common factor among cognitive abilities (the *G* factor) can (a) explain the variance in cognitive abilities through factor loadings represented in the black, solid arrows, and also (b) predict the variance in the factor loadings (i.e., CPEs) themselves through the gray, dashed arrows, with thus species of higher *G* exhibiting stronger factor loadings. The gray, dashed arrows will be referred to as cognitive integration coefficients throughout the present work, for brevity.

If *G* predicts the factor loadings positively as hypothesized, this finding would mean that although one high-order factor explains on average 58-68% of the variance among the cognitive abilities studied in the primate order (Fernandes et al., 2014; Reader et al., 2011), for species with comparatively low general intelligence the high-order factor can explain considerably less than that and thus abilities are more independent (less general). Conversely, for species with high general intelligence the high-order factor would be able to explain more than 58-68% and thus abilities are more generalized.

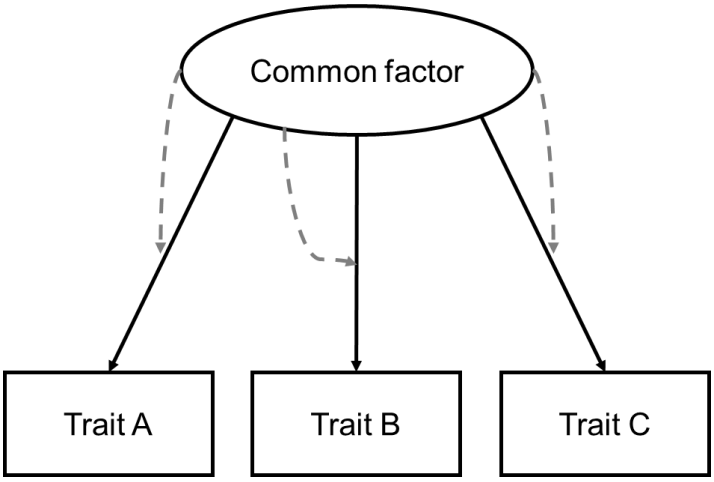

Figure 2.1. A hypothetical measurement model for a latent common factor among several observable traits, including factor loadings (i.e., part-whole correlations, represented in black solid arrows) and effects of the common factor upon the part-whole correlations (represented in gray, dashed arrows). The latter arrows are here termed "integration coefficients".

Note that the average of the CPEs for a given cognitive ability, taking all species in the sample, corresponds to the average factor loading of the *G* factor upon that cognitive ability in the sample[4]. These average factor loadings, one for each cognitive ability, will be reported as

[4] This procedure leads to identical factor loadings as would be produced by unit-weighted factor estimation. This is in fact a desirable technique to be used in the present sample (Figueredo et al., 1995 and Gorsuch, 1983). This is because factor loadings produced with Principal Axis Factoring and Principal Components Analysis in small samples are less reliable than those computed with unit-weighted factoring, considering that increasing standard errors of factor loadings with decreasing sample size prevents one from discriminating between the factor

well for reference. They should be identical to the already-reported average factor loadings from Fernandes and colleagues (2014).

2.2.2 Phylogenetic Controls

Species in a phylogenetic tree are not independent data points, but rather are related to some more closely than to others. A species' trait value can be a direct response to its immediate ecology (thus reflecting natural selection pressure), but can also be inherited from ancestor lineages. Disentangling these two sources of contribution to traits is essential, and controlling for phylogenetic inheritance from ancestors is important in order to test the hypothesis that a trait (or several traits) are a response to ecological pressures. With this in mind, the same analyses described above will be conducted a second time, this time controlling for phylogenetic inertia (that is, inheritance of traits rather than novel evolution) in all variables. To achieve this, I will estimate phylogenetic independent contrasts (see description below) for each variable, then all steps of the analyses described in Section 2.2.1 of the present Chapter description will be repeated. Tree data and statistical packages used will be the same as for Project 1.

The results will be reported alongside the non-phylogenetic analyses (i.e., those involving raw species data). It is important to note that correlative analyses involving phylogenetic controls do not *replace* other analyses and are not inherently more rigorous or correct. Instead, they answer different questions (Peñaherrera-Aguirre & Fernandes, 2018; Price, 1997; Thornhill & Fincher, 2013). Namely, correlative analyses without phylogenetic controls reveal whether there

---

loadings of the different variables appropriately (Bobko et al., 2007, Figueredo et al., 1995, Gorsuch, 1983; and Schmidt, 1971).

is an association between variables irrespective of whether the measures being correlated evolved at the same time in history; that is, they test simply whether one characteristic is present in taxa that exhibit other characteristics being studied. Phylogenetic comparative methods involving controls, on the other hand, reveal whether the variables in question have co-evolved *in tandem*; in other words, they reveal the degree to which they have been in lockstep through evolutionary time, covarying on the same phylogenetic tree branches across evolution. As such, phylogenetic controls permit examining whether the novel evolution of one characteristic is associated with the novel evolution of another. On the other hand, correlative analyses without phylogenetic considerations permit examining whether the maintenance of one characteristic is associated with the maintenance of another, even if the two were not originally selected simultaneously (Thornhill & Fincher, 2013). It may be that changes existed first in one trait, which permitted, in subsequent lineages, the evolution of the second trait, and this association between them was maintained as it was adaptive (e.g., the subsequent characteristic enhanced the functional property of the former characteristic acquired, but they did not arise in evolutionary lockstep). Such an association can be detected without phylogenetic methods, as the degree to which the origins of the two traits were tied *in time* is irrelevant in this case (Peñaherrera-Aguirre & Fernandes, 2018; Thornhill & Fincher, 2013).

Therefore, as both approaches reveal important information about data, results of both will be reported and discussed. Hypotheses forwarded in this study apply to raw species data analyses and phylogenetically controlled analyses; in other words, it is predicted that the effects of the common factor upon the strength of factor loadings were relevant both for the novel evolution of the integration of cognitive abilities and for its evolutionary maintenance. Descriptions of results for analyses that include phylogenetic controls will be prefixed with the term *evolutionary* or

postfixed with the term *in lockstep*, reflecting the notion that any association among traits that is identified using phylogenetic methods indicates an evolutionary association through time. Below, I detail the phylogenetic comparative method that will be used as a phylogenetic control:

Phylogenetic Independent Contrasts (Felsenstein, 1985): Although the last three decades saw the emergence of multiple phylogenetic comparative methods (Peñaherrera-Aguirre & Fernandes, 2018), employing phylogenetic independent contrasts (PIC) remains one of the most statistically conservative. Different from other procedures, this approach fixes the level of λ (a measure of how conserved, that is, preserved a trait is across evolutionary time; Nunn, 2011; Peñaherrera-Aguirre & Fernandes, 2018) of traits to its maximum (1.0). In simpler terms, this means that this approach assumes that closely-related species will be more similar to each other than more distantly-related species. PICs are then computed by estimating the difference between the values of a continuous trait of two species sharing a common ancestor, and this is repeated for all lineages in the phylogenetic tree, yielding as many contrasts as there are pairs of lineages. Since species with a more distant ancestor are expected to display greater difference in the trait in question, PIC takes into consideration the length of the branches in the phylogenetic tree (Nunn, 2011).

The estimation of PICs is illustrated in Figure 2.2. PICs are computed following a recursive pattern, beginning at the tip of the branches and going back in time through the phylogeny until the process reaches the root of the tree (Nunn, 2011). After each contrast is computed, the species that were involved in these calculations are excluded from the estimation of the subsequent contrasts. Consider three taxa, *Pan troglodytes* (B), *Pan paniscus* (C), and Homo sapiens (D). The first pair, B and C share a common ancestor (BC), whereas D shares a common ancestor with BC. While the first contrast is computed between B and C, the second contrast is

estimated between D and BC. It is worth noting however, since most analyses rely on data from extant species, PIC requires the reconstruction of hypothetical phenotypic values for the extinct common ancestors. The reconstructed estimate, in this case for BC, is not the average between species B and C, but alternatively, it is computed by weighing the values of B and C by the length of their corresponding branches. The second contrast is calculated as the difference between the reconstructed phenotype of BC with the observed phenotypic value of D.

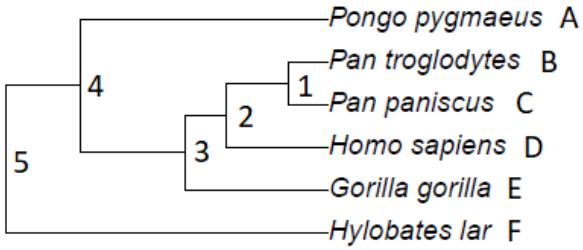

Figure 2.2. Simplified illustrative model containing only 6 species, representing where phylogenetic independent contrasts (numerals) lie in the phylogenetic tree, relative to measured species (i.e., tip) data (letters).

Note that phylogenetic independent contrasts for two variables are not necessarily orthogonal (i.e., uncorrelated and thus independent *of each other*). In fact, PICs are frequently used precisely to examine the strength of correlation among variables. PICs are only *independent* in the sense that, while original species data points were connected through a pattern of shared ancestry, once PICs are produced for a trait they can be used as data points that are no longer connected through ancestry (Felsenstein, 1985). In other words, PICs are independent of the phylogeny as if data had been residualized against it.

One of the advantages of employing the PIC method relative to other available phylogenetic methods is that a set of contrasts can be extracted as a dataset, and subsequent statistical analyses can be performed on the contrasts. For instance, the contrasts for trait X can be correlated to the contrasts for trait Y, to examine if the two traits have coevolved in lockstep through the phylogeny. Although several statistical procedures have phylogenetically informed homologue methods (e.g., path analyses and phylogenetic path analyses; generalized linear mixed models and phylogenetically generalized linear mixed model; meta analyses, and phylogenetic meta analyses; Peñaherrera-Aguirre & Fernandes, 2018), currently there is no phylogenetic equivalent for CPEM. However, extracting a set of PICs allows the computation of CPEs on the PICs after accounting for phylogeny. Put simply, a column of CPEs computed on PICs would reflect the evolutionary in-tandem correlation between two traits; but rather than leading to a single correlation coefficient for the entire sample like Pearson's *r*, using CPEs leads to a correlation coefficient for each lineage in the dataset.

To summarize: Entering species data for a trait into the phylogenetic tree permits producing phylogenetic independent contrasts for that trait, which are a column of data just as any other variable and as such can be used in subsequent analyses. Using PICs is equivalent to residualizing the trait against the phylogeny, thus eliminating the pattern of interdependence among the data points for that variable. Producing PICs for multiple traits separately permits subsequently entering them into correlative analyses just as can be done for any other variable. The correlations among traits, if the correlative analyses use PICs, are equivalent to partial correlation, where the effect of phylogeny is partialed out, whereas correlative analyses that purely use species data without any consideration of the phylogeny (e.g., without producing PICs) are equivalent to examining the zero-order correlations among traits.

3. Results

Results below are reported for the factor loadings of *G* upon the five cognitive abilities, and for cognitive integration coefficients for the five cognitive abilities. Recall that factor loadings, in this case, indicate how correlated performance on a cognitive ability is to the set of cognitive abilities using species data. If factor loadings for a given cognitive ability are estimated using phylogenetic independent contrasts rather than using species data, they are termed "evolutionary factor loading" and reflect the degree to which *G* and the cognitive ability have coevolved in tandem. If positive and strong, this evolutionary factor loading thus indicates that changes in the given cognitive ability occurred in lockstep across evolutionary time with changes in other cognitive abilities that comprise the *G* factor.

Recall also that the cognitive integration coefficient is the correlation of *G* values the factor loadings on a given cognitive ability. It is estimated across species, and if the coefficient is positive, species with higher *G* exhibit a stronger factor loading on that given cognitive ability. In turn, the evolutionary cognitive integration coefficient is the result of exactly the same procedure, but conducted on phylogenetic independent contrasts instead of on raw species data. If the evolutionary cognitive integration coefficient is positive, *G* and the factor loading for the given ability are tied across evolutionary time, changing in tandem and in the same direction through evolution.

3.1. Average Factor Loadings

While a continuous parameter estimate (CPE) for a given cognitive ability reflects its species-specific factor loading for that ability, the average of CPEs for that ability (i.e., taking all

species together) should be identical to factor loadings obtained by the unit-weighted factor scoring reported by Fernandes et al. (2014). Indeed as expected, identical average factor loadings were found by computing the average of CPEs for each cognitive ability, compared to the factor loadings reported by Fernandes et al.'s (2014) unit-weighted factor approach, and highly similar to those reported by Reader et al. (2011) which were produced by principal axis factoring. Results indicate strong and significant factor loadings for all five cognitive abilities, with loadings ranging from .684 to .904 without phylogenetic controls, and between .722 to .865 after accounting for phylogeny (see Table 2.1). In both sets of analyses tactical deception displayed the weakest factor loadings. Tool use and extractive foraging had the strongest loadings even after accounting for phylogeny, as also detailed (and discussed) by Fernandes et al., 2014.

3.2. Variations in Factor Loadings as a Function of G

The relationship of *G* with the factor loadings on each of the five cognitive abilities is reported below for each ability. Also, it is possible to estimate an "omnibus" relationship of *G* with the set of factor loadings; in other words, a general integration coefficient, reflecting whether *G* predicts how strongly the cognitive abilities are interrelated on average. This is operationalized as follows: The factor loadings for the five cognitive abilities are entered into a unit-weighted common factor (i.e., their standard scores are averaged into a single variable); then, the correlation of *G* with this common factor is estimated. The correlation index resulting from this procedure serves as an omnibus, protective estimate of the relationship of *G* with factor loadings in general. If its magnitude is moderate to high, it suggests that species with higher levels of *G* have more integrated abilities.

G positively predicted stronger factor loadings for all cognitive abilities, as shown in Table 2.1. Effects were of moderate to high magnitude, consistently for all abilities. The omnibus integration coefficient was .657 ($p < .05$).

Once phylogenetic contrasts were used, however, the magnitude of cognitive integration coefficients reduced somewhat, as is common in comparative studies (Price, 1997). It remained moderate and significant for tool use, extractive foraging, and innovation, but weak and non-significant for the two sociality-related cognitive abilities (social learning and tactical deception). Nevertheless, the omnibus cognitive integration coefficient remained significant and moderate in size (.283, $p < .05$), suggesting that *G* continues to predict stronger associations among cognitive abilities overall even when phylogenetic contrasts are used instead of species data.

Table 2.1.

Average factor loadings and cognitive integration indices for the five cognitive abilities, without and with phylogenetic control

Cognitive ability	**Factor loading**	**Evolutionary factor loading**	**Cognitive integration coefficient**	**Evolutionary cognitive integration coefficient**
Tool use	.905*	.846*	.636*	.412*
Extractive foraging	.879*	.865*	.632*	.349*
Innovation	.864*	.788*	.617*	.284*
Social learning	.784*	.809*	.653*	.113
Tactical deception	.688*	.722*	.604*	.153

Note: * $p < .05$.

A post-hoc hypothesis regarding the evolutionary cognitive integration coefficient involved the possibility that the cognitive abilities with a stronger evolutionary integration should be those that are, on average, more central to *G*. In other words, cognitive tasks on which *G* has a stronger factor loading should exhibit more integration as a function of increasing levels of *G*, because they are integral to the common factor. In fact the correlation between factor loadings and the magnitude of cognitive integration coefficients supported this hypothesis (r = .685). As for the analyses with species data, cognitive integration coefficients were largely similar (all between .604 and .653, as per Table 2.1), thus their size was not so strongly predicted by the factor loadings of the respective cognitive tasks (r = .396) although still positive.

Considering that size and significance of all coefficients for the species data analyses, and of three out of five coefficients for the analyses involving phylogenetic contrasts, post-doc analyses were then run to unpack the cognitive integration coefficients exhibited in Table 2.1. Each such cognitive integration coefficient in the previous table involved the effect of *G* upon the relationship of each cognitive task with *all* other cognitive tasks. This could be unpacked so that one can examine which pairs of cognitive tasks are most responsible for these coefficients. To illustrate: while the relationship of tool use with other cognitive tasks was demonstrated to increase as a function of *G*, it is possible that what increases as a function of *G* in that omnibus relationship of tool use with all other tasks is driven only by the relationship of tool use to one or two other particular tasks. Alternatively, it is possible that the relationship of tool use to each and every other task really increases as a function of *G*. This ought to be explored. To that effect, Table 2.2 displays the effect of *G* upon the *bivariate* relationship of each pair of cognitive tasks.

Table 2.2.

Predictive power of *G* upon bivariate correlations involving pairs of cognitive tasks; using species data (to the left of the diagonal) or phylogenetic contrasts (to the right of the diagonal)

Cognitive ability	**Tool use**	**Extractive foraging**	**Innovation**	**Social learning**	**Tactical deception**
Tool use	-	.189	.368*	.241*	.189
Extractive foraging	.613*	-	.327*	.136	.195
Innovation	.572*	.552*	-	-.007	.146
Social learning	.607*	.618*	.643*	-	.068
Tactical deception	.547*	.573*	.601*	.578*	-

Note: * $p < .05$.

Although *G* positively and strongly predicted bivariate correlations among all pairs of variables at present (i.e., using extant species data), it can be noted that, when using phylogenetic contrasts to examine the historical predictive power of *G* upon the in-tandem coevolution of traits, only three coefficients are significant. Of these three, two involve tool use and two involve innovation.

4. Discussion

This study aimed to examine the pattern of correlations among performance measures of various cognitive tasks, as a function of general intelligence (*G*), across primate taxa. It also aimed to examine if results differed in a phylogenetically-controlled design compared to a design

relying solely on species data (i.e., without a consideration of the degree of inter-relatedness of the primate species). These goals permit testing if the strength of the positive manifold *G* varies predictably across primates.

Data examination was guided by two evolutionary theory-derived hypotheses. Firstly, general intelligence should be an adaptation for organisms in generalist ecologies, which require not specialized cognitive phenotypes that permit dealing with a small set of adaptive problems, but rather should require cognitive flexibility to solve novel, complex problems and translate solutions across ecological subdomains. As such, it was expected that species with higher scores on the general intelligence factor (*G*) would exhibit stronger interrelations among the measures of cognitive performance (i.e. high factor loadings for the tasks), rather than a disintegration which would be indicative of specialization.

Secondly, it was hypothesized that, as the adaptive usefulness of *G* would *require* integration among cognitive domains, supportive results for the aforementioned prediction should be found both without and with controls for phylogenetic non-independence. Such controls permit examining not only whether traits are associated presently, but also through evolutionary time (i.e., whether changes across traits happened in lockstep in evolutionary history).

Overall, results supported both hypotheses. The cognitive integration coefficient (representing the correlation between *G* and the factor loading for a given measure, as indicated in Figure 2.1) was moderate to high in size for all measures without phylogenetic controls, and moderate for three out of five measures after phylogenetic controls. The overall protective test was also moderate in size after the phylogenetic controls. Interestingly, the analysis employing the method of correlated vectors to examine the post-hoc possibility that the cognitive integration

coefficient was stronger for more *G*-central measures (i.e., those with a stronger average factor loading) led to a confirmation of this hypothesis. In other words, cognitive abilities that exhibit a high average factor loading in the primate order exhibit a stronger integration with other cognitive abilities as a function of *G*.

4.1 Tool Use and Innovation as Drivers of Cognitive Integration

Post-hoc analyses indicated that, while all bivariate correlations among all cognitive tasks increase as a function of *G* in this sample of primate species, once *evolutionary* bivariate correlations are considered (i.e., using phylogenetic contrasts), only a few were significant and still moderate in size. Two of these involved tool use and two involved innovation. This means that the increases in correlations among cognitive tasks occurred in tandem with increases in *G*, whereas increases in correlations among other tasks occurred in different tempos (either before or, most likely, lagged behind).

These findings reinforce the notion that tool use and innovation were central drivers of the evolution of intelligence in primates (Fernandes et al., 2014; Reader & Laland, 2002). While, conceptually, innovation is indispensable to *novel* problem-solving, tool use permits control and exploration of the ecology (Darwin, 1871; Gibson & Ingold, 1993; Washburn, 1959; Wynn, 1988). Moreover, the centrality of tool use in the evolution of general intelligence may relate to the finding that the pattern of evolution of the cerebellum in primates is moderately commensurable to that observed for *G* (see Chapter 1), as activity in the cerebellum appears to be an important substrate for the coordination of motor skill necessary for the implementation of cognitive problem-solving (Doyon et al., 2003; Salman, 2002). This body of work reinforces the notion that tool use is an important mechanism for promoting the efficient exploitation of a

diverse, complex habitat in a variety of ways as unique opportunities arise (Parker, 1978), including novel ways due to the capacity for innovation.

The present findings also agree with Parker and Gibson's (1977, 1979) distinction between intelligent and context-specific tool use. They have proposed that, in order for tool use to qualify as "intelligent" it needs to be flexibly adjustable to novel situations in innovative ways. While tool use in the sample used in the analyses in the present study is exhibited by many species, it is an integral part of general intelligence in some taxa only. That is, it is highly correlated with other cognitive skills, especially with innovation, in only part of the sample. This tool use-innovation complex is, not surprisingly, supported by the evolutionary, lockstep correlation between these two variables as a function of increasing *G*.

In line with the previous identification that sociality-related measures (social learning and tactical deception) exhibit the lowest factor loadings from *G*, slowest evolutionary rates and strongest conservatism among the five cognitive abilities used in this study (Fernandes et al., 2014), here they were the only two abilities that failed to exhibit stronger factor loadings on the *G* factor as a function of level of *G*, in the phylogenetically-informed analyses. This suggests that they may be less central to *G* compared to other aspects of cognition. Nevertheless, these results are far from indicative that social cognition is independent from *G* – they still exhibit high factor loadings from *G*, less conservatism and much faster evolutionary rates than most traits in primates (even other behavioral traits; Blomberg et al., 2003; Kamilar & Cooper, 2013). Moreover, even though cognitive integration coefficients were not significant for social cognitive abilities after use of phylogenetic contrasts, they were high and significant when using species data. This suggests that while the integration of social cognition may not have coevolved as much in lockstep with so-called technical or physical cognitive abilities across time, they do

exhibit an association that has been maintained. Furthermore, social transmission has been implicated as an important potentiator of innovation rate across animals (Arbilly & Laland, 2017).

4.2. The Evolution of Cognitive Integration

In primate studies, it is common to find descriptions of capacity for the comprehension of generalized object identity; categorization of abstract relations, and thus the capacity to interchangeably rely on different entities within a category to flexibly solve problems (de Waal, 2016). These capacities suggest that many primates are capable of generalization from one domain of adaptive problems to another when problem solving, to translate applications. In other words, cognitive domains appear to *interact*. The present study suggests that some primate species exhibit a stronger correlation among their cognitive abilities. While this correlation does not necessarily imply that such species are able to *translate* applications of one cognitive skill to others, it would be difficult and unparsimonious to defend that the evolution of general intelligence and of stronger intercorrelations among its components would involve the opposite possibility: that of impenetrability and lack of communication, transfer, or interaction among domains of cognition.

The notion that ecological generalism is tied to cognitive complexity is longstanding (Parker, 1978). The study of the so-called proactive-reactive axis of behavior has also converged with this notion for more than 50 years. Proactivity and creativity in problem-solving are exhibited in exploratory behavior and in response to environmental changes, and proactive organisms use behavior to control and manipulate environmental conditions. Proactivity and creativity appear to be associated with ecological generalism in comparative studies of primate

species (Glickman & Sroges, 1966; Visalberghi & Mason, 1983; see also Mettke-Hofmann, 2014 for a review). More recently, it is becoming clearer that generalists are also better learners, and have evolved more connectivity in the brain (Mettke-Hofmann, 2014), along with more innovation and better performance on complex problem-solving cognitive abilities (Henke-von der Malsburg & Fichtel, 2018). The present study contributes to this long, albeit relatively sparse body of knowledge by adding a piece to the puzzle, as not only there is evidence of links between ecological generalism, complexification of the neuroanatomy, and performance on complex and creative problem solving, but the present study also adds evidence that overall cognitive performance has coevolved with the integration among cognitive abilities in the primate order. Arguably, this integration permits the use cognitive skills interactively.

This growing body of knowledge can also offer insight to our understanding of the evolution of cognition in the human lineage: Several primate lineages, including hominins, were able to adapt to largely fluctuating environmental conditions in the late Tertiary. They are frequently proposed to have been generalists and opportunists (see Parker, 1978; Ripley, 1984). Reader et al. (2011) demonstrate that increased levels of general intelligence are also comparatively recent in the primate phylogeny, in largely the same lineages. The present results suggest that increasing general intelligence predicts increasing cognitive integration. Note that the correlation of *G* with the cognitive integration coefficients was observed both for species data and for phylogenetically-controlled analyses, which permit inferring that they have covaried *in tandem* across phylogenetic time. This implies that high levels of cognitive integration are a comparatively novel evolutionary feature. Moreover, Chapter 1 examined degree of change in the evolutionary rates for *G* in primates, using the same dataset compiled by Reader et al. (2011), and identified a strong acceleration pattern from the root to the tip of the phylogenetic tree. This

compounds the interpretation that high scores on *G* in several primate taxa are a considerably recent phenomenon, further solidifying the hypothesis that the evolution of integration of cognitive abilities should also be recent.

It should be noted that the hypothesis that ecological generalism is behind the evolution of *G* and of stronger associations among cognitive skills is currently being tested (Peñaherrera-Aguirre & Fernandes, in preparation). Preliminary analyses, both with species data and with phylogenetically-informed methods, indicate that habitat generalism (i.e., the number and diversity of habitats a species occupies) is predictive of general intelligence and of the intercorrelations among the five cognitive skills also used in the present study. While the correlative design used does not permit pinpointing empirically whether habitat generalism is a cause or an effect of cognitive evolution, it suggests that they have coevolved in lockstep in the primate phylogeny, being co-adapted, as hypothesized in the present work.

4.3. Implications for the Understanding of Specialized Abilities

Framing the results in terms of increased integration among cognitive abilities as a function of *G* does not make explicit the implications to the other end of the *G* continuum: species with a lower *G* average exhibit more independent variation in their performance on cognitive tasks. To illustrate, some of the species with a low *G* score may excel in tool use but exhibit little of other cognitive abilities; another species may frequently execute extractive foraging but exhibit little social intelligence or use of tools. As such, there appear to be many alternative ways of being a low-*G* species, and due to the low covariation among performance on cognitive tasks in these species it is unlikely that any species performs poorly on *all* cognitive tasks. In other words, the present results suggest that in species with lower *G*, cognition is more

specialized and there is thus more diversity among them. Potentially this involves less communication and interaction among cognitive adaptations, although the neuropsychology behind the present results is still not explored.

Considering this, contrary to concerns expressed by some researchers (e.g., Hodos & Campbell, 1969; Tomasello & Call, 1997), a discussion of *G* does not reinforce the antiquated concept of a *scala naturae*. Instead of suggesting that species with higher *G* are better adapted than others or superior in performance *across domains*, the present line of research and the present results suggest that species of low *G* exhibit a different pattern of interrelations among cognitive abilities, rather than having a cohesively low performance on all of them. In other terms, they are possibly low on the general factor simply because they do not require such an integrated factor for their ecology; instead, they exhibit specialized abilities and may excel in them, without requiring commensurable and interactive action in other domains of cognitive performance.

Just as the present study employs the hypothesis that high *G* has evolved in ecological generalists, conversely those with lower *G* scores are expected to have evolved as ecological specialists. Through the lens of the generalism-specialism continuum, it is also impossible to reinforce a *scala naturae* argument with the present results: Specialization and generalization each have their particular advantages and disadvantages, in terms of evolutionary adaptedness, one not being intrinsically superior to the other (Parker, 1978). While generalism is advantageous when flexibility to cope with varying conditions is necessary, specialism is advantageous in more stable conditions where a few resources can be confidently and thus more efficiently used.

4.4. Contrast to the Spearman's Law of Diminishing Returns

In human samples of individuals, those with higher IQ scores tend to exhibit lower factor loadings on the general factor of intelligence, suggesting that cognitive performance is more specialized in high-performing individuals. This phenomenon Spearman termed "law of diminishing returns" (Spearman, 1927; see also Jensen, 2003). At first impression, the results described in the present study go counter to that well established phenomenon.

Few evolutionary accounts of the law of diminishing returns have been proposed and only as post-hoc explanations (for a review, see Woodley, 2011), with most of the literature being atheoretical instead. Evolutionary accounts propose, for instance, that (a) pleiotropic deleterious mutations may have an effect across a wide variety of brain structures and processes, thus creating an apparent integration among low levels of ability in those with low IQ (i.e., jeopardizing efficiency across the brain); or that (b) cognitive differentiation among those with high IQ scores is best explained by individual differences in life history speed, while recognizing that contrary to cross-species and cross-population studies, individual differences in general intelligence and in life history speed paradoxically do not correlate.

Like several relationships among variables that conflict in direction or strength when comparing within-species to between-species results (e.g., Rushton's paradox of the relationship between intelligence and life history; Figueredo et al., 2017), Spearman's law of diminishing returns among humans presents a paradox as it goes counter the results of the present study. It is clear that these proposed post-hoc explanations do not clash with the present theoretical formulation of the relationship of ecological generalism to the factor loadings; that is, they are not simply the reverse of the hypotheses forwarded in the present study. No hypothesis has been put forward, let alone supported, that the strength of the general factor of intelligence reduces

due to ecological pressures for generalism. Having said that, it is imperative that future studies make a comprehensive and simultaneous *competitive* examination of the effects of the need for ecological generalism, life history variation, and pleiotropic mutations affecting brain functioning to compare their putative effects upon factor loadings on general intelligence. That way, all hypotheses regarding variations in the strength of general intelligence factor loadings are directly compared. This is, however, not possible with existing data on primates given the considerably small overlap between datasets containing the aforementioned variables, and scarcity of data on the effect of mutations on a variety of species. Furthermore, such an empirical comparison of hypothesis would require using only one level of analysis (either cross-species or cross-individuals), whereas the present work is conducted at the cross-species level but the law of diminishing returns has been proposed for individual differences only.

4.5. Limitations and Future Directions

It must also be noted that the ethological variables used in this study were counts of behavior, rather than fine-grained measures of the degree of problem-solving success in the implementation of the behaviors. Still, previous conclusions drawn from studies using this same database largely converged with those derived from laboratory cognitive challenges (for a review, see Reader et al., 2011), and it must be noted that error is reduced when using pooled databases that integrate multiple sources such as the database used in the present study (Lubinski & Humphreys, 1996).

As described in the Method Section of this Chapter, the use of phylogenetic contrasts, rather than fitting the phylogenetic tree topologies to the particular phylogenetic signal of each measure, is a conservative approach and thus may lead to underestimates of correlations. As

such, the use of contrasts provides a strong test of the hypotheses, and protects against the possibility that the phylogenetic signal of the measures was lower than 1.0 (which is equivalent to using contrasts) in case of measurement error in the ethological observations.

The possibility that the coefficients reported when using contrasts were underestimates is also likely if we consider that phylogenetic controls partly eliminate the effects of ecology as well, not only of phylogeny (Fernandes, 2014; Price, 1997)[5]. Closely-related lineages are phenotypically similar not only because they share a common ancestor (which is the premise of phylogenetic controls), but also because their niches tend to share many characteristics. Once there is speciation, the daughter lineages do not suddenly occupy niches that are the polar opposites of each other, nor randomly-distributed niches. Rather, their niches tend to have more similarities than those of distantly-related species. In line with this, in a study of phylogenetic signal across carnivoran species, all environment measures studied (local temperature,

[5]It is also rarely pointed out that inertia in a trait exists in a lineage only if there is *selection* for the maintenance of the characteristic. In other words, if a trait confers selective advantage, it is maintained if already widespread in the lineage. Traits inherited from ancestor lineages do not *need* to undergo inertia. Cognitive and behavioral traits, especially, are continuous polygenic characteristics influenced by vastly numerous loci upon which selection can act, modifying the trait qualitatively or quantitatively (Plomin, Owen, & McGuffin, 1994). Such modifications can take place in few generations if important changes in ecological landscapes take place in speciation (Gangestad, 2011). As such, the argument that species that diverged phylogenetically millions of years ago should retain similarity simple because of shared ancestry (implying that only phylogenetically-controlled analyses are valid, or even that they do not contain biases) is hard to defend.

precipitation, evotranspiration, etc.) were moderately and significantly phylogenetically conserved (Fernandes, 2014). As such, controlling for phylogenetic non-independence through the use of contrasts or any other technique partially eliminates true adaptive variance due to the control for niche similarity. This undesirable effect is not particular to contrasts specifically, and thus still unavoidable when one wishes to examine the degree to which traits have covaried in lockstep through evolutionary history no matter what the phylogenetic technique employed is.

It is also important to note that, although the present results stress a statistical integration among cognitive performance measures, it does not demonstrate or assume that integration among neuroanatomical areas, neural networks, or any subjacent physiological set of processes parallels the results exhibited. Even though differences among species in some neurological substrates must exist behind the reported cognitive integration, it is unclear what the species differences are: whether they lie in the correlations among sizes of brain areas, or in the connectivity among areas, or in diffusion of cognitive functions into less regional specificity. A stronger correlation among performance scores on different cognitive problems can potentially be explained by multiple different neuropsychological hypotheses. The present findings are not evidence for any particular alternative and cannot be used to discriminate among them. Considering the profuse reliance on volumetric measures in the primatological and anthropological literature on intelligence and the arguably pressing need to test the validity of this approach, Chapter 3 will have the goal of examining the first possible explanation outlined above for the findings exhibited in the present study: that brain structures are correlated in size more strongly in some taxa.

Chapter 3. Examining the Magnitude of Correlations among Sizes of Brain Regions across the Primate Phylogeny: Stable Magnitudes across Lineages of Different Brain Sizes

1. Introduction

As in the case of relations among cognitive abilities, the statistical relations among sizes of brain regions across species have been explored in a simple way only. The possibility that the strength of the relations among sizes of brain regions is stronger in clades of larger brains has not been examined. In other words, no hypotheses have been tested, in primates or other clades, regarding the possibility that the correlation among sizes of brain regions increases or decreases as a function of species brain size. While the finding that performance scores on cognitive abilities become more correlated among themselves in species of higher general intelligence (*G*) may suggest that this is the product of integration among brain regions, this is a large assumption that should not be made without examination. Furthermore, if brain regions are indeed more integrated, in what sense are they so? In terms of stronger correlations among sizes of regions, or in having more projections connecting them, or in other ways that do not necessarily involve neuroanatomical volumes? This is an empirical question that must be examined.

This study has the goal of testing specifically the first possibility, namely that brain structures exhibit stronger associations, in species of larger brains, in terms of correlations among their volumes. The use of volume measures is widespread in the comparative literature, where they are taken as substrates for intelligence (as detailed in Chapter 1; for further reviews, see Deaner et al., 2000; Deaner et al., 2007) and they are even employed as proxies for it in the absence of robust ethological or laboratory measures of cognitive performance (Shultz & Dunbar, 2010). Chapter 1 examined the evolutionary rates and selection patterns of general

intelligence and several commonly-used neuroanatomical volume measures in primates, demonstrating that comparability is overall limited. In the present Chapter, I will present a test of possible increases in correlations among neuroanatomical volume measures as a function of brain size, which will serve as comparison test to the increases in correlations among cognitive abilities as a function of general intelligence documented in Chapter 2. If neuroanatomical volume measures do serve as good substrates for intelligence to the point of serving as proxies (e.g., Shultz & Dunbar, 2010), the highly consistent integration observed among cognitive performance measures should replicate when the same analytical approach in neuroanatomical volume measures.

1.1. Concerted or Mosaic Volumetric Evolution of Brain Structures?

Multiple studies in the past several decades have attempted to examine whether, when comparing species of different brain sizes, the volumes of brain regions have evolved separately or in a coordinated manner. In other words, it is of interest to understand if brain regions have tended to vary in size *proportionately to each other*, or alternatively if in some taxa changes in the volume of brain areas were largely independent from changes in other areas (frequently called mosaic evolution; Gould, 1977)[6]. Conclusions drawn from these studies are conflicting, as

[6] Following the tradition of the field for consistency, the terms "mosaic" and "concerted" will be used throughout the text, respectively referring simply to (a) a lack of or low correlation between the sizes of various brain regions, when comparing species; and (b) the presence of correlations among the sizes brain regions, when comparing species. For instance, if regions A, B, and C have sizes 1, 2, and 3 in species *x*, and sizes 2, 4, 6 in species *y*, brain regions vary in a perfectly concerted manner in this dataset of species. When using these terms, no implications are made

some indicate that brain regions increase and decrease in size largely in a coordinated fashion (e.g., Finlay et al., 2001; Yopak et al., 2010), whereas others pinpoint specific regions that exhibit independent evolution in terms of size (e.g., Hager et al., 2012; Smaers & Soligo, 2013), although some recent essays propose that *some* regions or systems evolve in a concerted manner among themselves, but these sets of coordinated regions vary in size independently from other brain regions (Barton & Harvey, 2000; Smaers et al., 2011; Whiting & Barton, 2003). For instance, the size of the neocortex in primates is argued to have coevolved with the size of the cerebellum and with that of the diencephalon, but independently of the mesencephalon and the medulla (Barton & Harvey, 2000). A hybrid model between mosaic and coordinated evolution of brain structures is thus sometimes proposed (Barton, 2001).

Such a hybrid model may introduce more detail and present a less extreme alternative for the explanation of neuroanatomical evolution. However, the state of this debate remains confusing. Some authors have attempted to organize the field by examining whether perhaps some clusters of taxa exhibit mosaic evolution in the brain, whereas in other clades brain regions correlate highly in their sizes (Barton & Harvey, 2000; Clark et al., 2001; Frahm et al., 1982; Smaers & Soligo, 2013). While informative, two issues arise with this clade-specific analysis:

1) Demonstrating that the correlations of sizes of brain regions in one clade are different from the correlations observed in another clade tends to simply explain away the question of

---

about evolutionary interpretations or about whether the sizes regions have evolved in lockstep; they are instead purely descriptive terms and any evolutionary hypotheses must be tested. Likewise, no assumptions are made about whether brain regions that exhibit concerted change are more connected, functionally interactive, or complementary in their purpose – these possibilities should be examined empirically.

what accounts for correlated versus independent neuroanatomical evolution. It does not elucidate what may be behind correlations among sizes of brain regions, rather it only shows that *for some reason* still unexplained, clades may differ to a certain degree in the strength of these correlations. As such, clade-specific analysis is a descriptive but not explanatory approach.

2) Comparing a set of species of larger brain size to a set of species of smaller brain size as an effort to examine if brain size accounts for whether there has been independent (i.e., mosaic) volumetric evolution of brain regions or correlated evolution is an approach that leads to little sensitivity in the discrimination among brain sizes. It is akin to the common median split approach in individual differences studies: the *n* is reduced to half within each of the two analysis conducted (or to a third or a fourth if three or four sets of species are compared), and then comparisons among sets of species rely upon an *n* of two (or three or four) clades. This frequently leads to simply qualitative comparisons among sets of species, because an *n* of two is evidently not sufficient for any informed statistical comparison. As in Chapter 2, the continuous parameter estimation model addresses these statistical concerns by permitting fine grained analyses without splitting the sample into clusters, while still permitting analyses to be phylogenetically informed.

Brain regions are not constrained genetically to covary in size; covariations, when they exist, arise due to selection favoring them, as the concerted action of multiple brain regions appears adaptive. This is evidenced by the identification of largely independent genetic loci for variation in different key parts of the brain (Hager et al., 2012). Therefore adaptationist-informed hypotheses of stronger or weaker covariation among brain regions in terms of their sizes can be generated and tested.

1.2. Objectives and Hypotheses

A realization that needs to be considered and motivates the present work is that a correlation coefficient for the association between the volumes of two neuroanatomical regions permits only estimating the *average* association between variables in a given sample. It does not lend itself to examining whether variables are more associated in a certain part of the sample than in others; that is, other techniques are necessary to examine if the correlation strength is not constant across subsamples (e.g., certain species in a dataset of primate species). If, among taxa of larger brains, brain components start varying in size in a more interlocked fashion (i.e., in coordinated) or alternatively in a looser (i.e., more independent) fashion compared to way that occurs in taxa of smaller brains, a simple correlation coefficient as described above will not be able to detect this phenomenon. This will only be perceptible if one examines whether the *strength* of the correlation coefficient itself varies as a function of brain size (or as a function of any variable of interest).

As demonstrated in Chapter 2 that cognitive abilities become more integrated (i.e., correlated) in taxa of higher general intelligence, it is difficult to imagine how this would not lead to questions about whether a similar phenomenon takes place at the neuroanatomical level. In other words, does the integration among abilities necessitate an integration among brain regions? No single hypothesis is proposed for this study, considering that correlations among sizes of brain regions are not the only possible avenue for explaining the already detailed increased correlations among cognitive abilities in species of high average intelligence (Study II). Instead, there are multiple working hypotheses:

1) Larger brains would exhibit greater inter-region integration; that is, the correlations among volumes of brain regions are larger than in small-brained species. This hypothesis is based on the premise that the general intelligence observed in species of large brains coopts activity from multiple brain regions. As such, it would be unlikely that some regions would evolve considerably in size in an independent fashion relative to other regions, and it could thus be expected that volumes of regions would exhibit high factor loadings from a common factor in a comparative analysis.

2) Larger brains would not exhibit greater inter-region integration at the broad level (i.e., when analyzing the main fundamental brain parts), but rather inter-region integration would be observed among components *within* one fundamental brain part. This hypothesis rests on the premise that it is not the whole brain, but rather one fundamental part of it that is largely responsible for general intelligence. As such, other brain regions outside this fundamental region may vary in volume more freely, but subregions within this fundamental region would be largely integrated (i.e., their volumes would exhibit high factor loadings from a common factor in a comparative analysis). Specifically a hypothesis is here forwarded that the subregions of the telencephalon may be more strongly correlated in species with a larger telencephalon, as a) the telencephalon has long been recognized as an important target of selection for increases in primates, especially in more intelligent species (Stephan & Andy, 1964); b) it is the focus of most hypotheses of cognitive evolution (particularly in the neocortex – having even been described as ‘the crowning achievement of evolution’ (Rakic, 2009) – and the hippocampus; Shultz & Dunbar, 2010), c) somewhat detailed data on subregions exist for numerous primate species, but not for other major regions of the brain (Stephan et al., 1981).

3) No relevant difference would be observed when comparing species of larger brains to others in the factor loadings of brain regions. As volume measures were not commensurate with general intelligence in their evolutionary rates and modes of selection pattern (see Chapter 1; i.e., they do not appear a sufficient substrate for it, evolutionarily), volume measures of brain regions may also not be more integrated in larger brains as a substrate for general intelligence. In other words, the integration findings of Chapter 2 may not replicate for volumetric measures of neuroanatomy in the present Chapter, just as selection regimes for volumetric measures of neuroanatomy are disparate from those for general intelligence.

2. Method

Analyses will be done at two levels of brain structure division:

1) An analysis conducted done estimating the predictive effect of total brain volume upon the common factor loadings of a high-order latent construct comprised of the volumes of its fundamental parts, including volume measures for the pons, medulla oblongata, cerebellum, mesencephalon, diencephalon, and telencephalon. These structures correspond to the initial, main subdivisions of the embryonic primate (and vertebrate, in general) brain.

2) A second analysis will be conducted estimating the predictive effect of telencephalon volume upon the common factor loadings of a lower-order latent construct comprised of the volumes of its components. Specifically, components to be included are: The neocortex (i.e., isocortical grey and the underlying white matter), hippocampus, schizocortex(i.e., entorhinal, perirhinal and presubicular cortices), striatum, septum, lobus piriformis, bulbus olfactorius, and bulbus olfactorius accessorius.

2.1. Measures

For the first set of analyses, regarding the brain and its fundamental, high-order components, volumetric data will be compiled for 67 primate species for all variables. Data sources are the database developed by Stephan and colleagus (1981) with complementary data from Matano and colleagues (1985) for pons volumes, and the recent dataset with data for the same brain components for additional species and for some of the same species (i.e., measurements on new specimens) by Navarrete and colleagues (2018). Averages for each species will be computed when data are taken from multiple sources.

For the second set of analyses, regarding the telencephalon and its components, volumetric data will be compiled for 45 primate species for all variables, as fewer species have specific published details on the components of the telencephalon. The data source will be the database developed by Stephan and colleagues (1981), with complementary data (i.e., measurements on additional specimens) for some of the same species only for the neocortex, hippocampus, and striatum coming from Navarrete and colleagues (2018). Similarly, averages will be computed for each species, when data exist come from the two sources.

These datasets exhibit advantages and limitations. Their main apparent limitation is that volume measurement, for most species, relied on few specimens, even in the case of the recent database generated by Navarrete and colleagues (2018). However, considering the paucity of alternative measurements, using these data permit comparison with other comparative studies, especially in the case of the Stephan and colleagues' (1981) dataset, as it has for long been the main source for comparative analyses using neuroanatomical volume measures in primates (e.g., Deaner et al., 2007; Dunbar, 1992; Lindenfors et al., 2007; Sandel et al., 2016; Shultz & Dunbar,

2010). The inclusion of the recent, complementary data by Navarrete and colleagues permit reliance upon a somewhat larger number of species and specimens and thus higher confidence in parameter estimates.

All measures will be log-transformed for all analyses, because of the well-known skewness in comparative data (Harvey, 1982). Considering the well-known allometric relations of brain structures with body size (Montgomery et al., 2016) but also concerns that residualization of volumetric measures against body size eliminates truly adaptive variance from brain regions (Fleagle, 1985; Jeschke & Kokko, 2009; Roff, 2011; Smith, 1980), analyses will be presented both for residualized and non-residualized (i.e., absolute) measures, with no single preferred approach. These two operationalizations of neuroanatomical volume measures permit examining (a) if their absolute sizes are more strongly correlated in species of larger brains, and (b) if their residual sizes net of body size are more strongly correlated in species of larger residual brain size, residualized against body size as well.

2.2. Analyses

The continuous parameter estimation model (CPEM; Gorsuch, 2005) will be employed to examine, at the first level of analysis, if overall brain size predicts the strength of factor loadings of the principal brain regions. In the second level of analysis, CPEM will be used to examine if telencephalon volume predicts the strength of factor loadings of its subregions. Recall that a factor loading represents a part-whole correlation whereby, in this study, the *part* refers to the volume of a specific neuroanatomical region, and the *whole* represents the set (i.e., aggregate) of all regions in the brain, or in the case of the second level of analysis, the set of all subregions in the telencephalon.

As defined in Chapter 2, CPEM produces continuous parameter estimates (CPEs, i.e., factor loadings) for each variable *for each species* by computing the product of its *z*-score and the *z*-score of the common factor (i.e., the aggregate or *whole* described above). Each species thus has a factor loading estimate, for each neuroanatomical region in the dataset. In the dataset, a column of the factor loadings for a neuroanatomical region (i.e., a list of the factor loading values, one by species, for a given neuroanatomical region) will be used for two purposes in this study, exactly as was done in Chapter 2:

1) The values in the column will be averaged, to identify that average factor loading for that neuroanatomical region (i.e., the average part-whole correlation, where part is the region and whole is the set of regions). The average is exactly equivalent to the factor loading that would be identified if Unit Weighted Factoring were used, and very similar to the factor loading that would be identified if Principal Axis Factoring were used (see Chapter 2). Thus, were one to conduct a simple exploratory factor analysis using the present database of neuroanatomical volumes using data for primate species, the factor loadings identified in it represent simply the *average* factor loadings for the sample of species.

2) That column can be correlated to a column of brain size for the respective species (or in the case of analyses involving only the telencephalon, the column can be correlated to a column of telencephalon size). This yields the "neuroanatomical integration coefficient"[7], akin to the "cognitive integration coefficient" detailed in Chapter 2.

[7] Recall that the term "integration" is used without implying that regions are more strongly connected, share functions, or are connected in any other sense; rather, it refers to stronger correlations among neuroanatomical regions in terms of their sizes.

If the neuroanatomical integration coefficient is positive and moderate or strong, meaning that brain volume is positively predictive of factor loadings, it follows that in some taxa the factor loadings are more positive, whereas in others they are less strongly positive, or even somewhat negative. This potential integration would mean that, for some lineages, estimated factor loadings would be higher than the average factor loadings identified in (1), and for others they would be weaker or reversed in sign. This does not mean, however, that the average factor loadings estimated with a simple exploratory factor analysis are completely uninformative if an integration hypothesis is confirmed. As with any variable they would be informative of the central tendency for the sample, just not necessarily fully representative of the whole range of strengths of factor loadings observed.

As in Chapter 2, all analyses will be conducted with and without phylogenetic controls, and compared. Analyses purely using species data (without taking into account their phylogenetic non-independence) permits assessing the current correlation among variables, but not whether they have coevolved *in lockstep* through evolutionary time. They may have evolved at different moments in evolutionary history, and all this analysis suggests is that there has been selection for the maintenance of the associations (Peñaherrera-Aguirre & Fernandes, 2018; Price, 1997; Thornhill & Fincher, 2013). Analyses taking into consideration the phylogenetic non-independence among the species, on the other hand, permit examining whether traits have changed in tandem through evolutionary history (i.e. covarying on the same phylogenetic tree branches through time). Descriptions of results for analyses that include phylogenetic controls will be prefixed with the term *evolutionary* or suffixed with the term *in lockstep*, in order to reflect the notion that associations among traits that are identified using phylogenetic approaches indicate an association in tandem through evolutionary time.

To summarize: Results will be presented for (a) the average factor loading of the common factor of brain regions upon each individual brain region, (b) the relationship of brain size with the set of species-specific factor loadings of each individual brain region. (a) and (b) will also be reported for analyses using phylogenetic independent contrasts, respectively indicating the average evolutionary factor loading for each brain region, and the evolutionary anatomical integration coefficient. These four sets of results (a and b, without and with phylogenetic control) will first be reported with variables not having been residualized against body size, then they will also be also be reported with all variables having been residualized against body size.

Exactly the same set of reports will then be exhibited for analyses involving telencephalon size and its subregions, after reports of analyses involving brain size and its regions.

3. Results

3.1. Average Factor Loadings and Anatomical Integration Coefficients for Brain Structures

Examining the correlations among the main brain structures, it is evident that a single common factor accounts for their covariance (see Table 3.1). This is observable both without and with phylogenetic controls, indicating that they are strongly correlated across evolutionary time. However, no changes in the factor loadings occur as a function of overall brain size, as anatomical integration coefficients (reflecting the correlation of brain size with factor loadings for a given brain region) for all regions exhibited low and non-significant magnitudes (Table 3.1). The omnibus anatomical integration coefficient (operationalized as the correlation between brain size and the standardized average of all CPEs for all brain regions) was .015 ($p > .05$), and the omnibus evolutionary coefficient was .131 ($p > .05$), compounding the assertion that no integration of brain regions occurs as a function of brain size. Moreover, the correlation between

the vectors of average factor loading and anatomical integration coefficient was weak and non-significant both for the case of raw and evolutionary coefficients: r = -.146 and .298 (p> .05) respectively.

Table 3.1.

Average factor loadings and integration indices for regions of the primate brain, without and with phylogenetic control

Brain region	**Factor loading**	**Evolutionary factor loading**	**Anatomical integration coefficient**	**Evolutionary anatomical integration coefficient**
Telencephalon	.993*	.982*	.022	.141
Diencephalon	.998*	.990*	-.018	.137
Mesencephalon	.990*	.921*	.010	.157
Cerebellum	.994*	.963*	.072	.205
Medulla	.992*	.919*	-.014	-.009

Note: * $p < .05$.

Table 3.2 displays equivalent analyses to those reported in Table 3.1, but with each brain region's volume being residualized against body size. While average factor loadings largely replicate the above reported results, anatomical integration coefficients exhibit differences: mesencephalic and cerebellar sizes appear more strongly correlated to other brain regions as a function of brain size. A non-significant but moderate correlation was found between the vector of factor loading and the vector of anatomical integration coefficients reported in Table 3.1 (r =

.474, p > .05), suggesting that it is possible that the reported integration effects are especially stronger for regions that are on average more central to the common factor. The omnibus anatomical integration coefficient (i.e., the overall effect of brain size upon factor loadings in general) was not of negligible size (r = .200) but failed to reach significance (p = .10).

Once phylogenetic controls were employed, positive integration results for the few regions that did exhibit integration before did not replicate, suggesting that brain size evolved at a different pace compared to the volumetric integration of the cerebellum and the mesencephalon with other regions. No evidence was found that regions that are on average more central to the common factor (i.e., having stronger average factor loadings) exhibited stronger anatomical integration coefficients (r = -.416, p > .05). The omnibus anatomical integration coefficients was -.058 (p > .05), compounding the observation that no consistent changes occur in the strength of factor loadings of brain regions as a function of brain size.

Table 3.2.

Average factor loadings and integration indices for regions of the primate brain residualized against body size, without and with phylogenetic control

Brain region	**Factor loading**	**Evolutionary factor loading**	**Anatomical integration coefficient**	**Evolutionary anatomical integration coefficient**
Telencephalon	.905*	.945*	.136	.010
Diencephalon	.959*	.961*	.205	-.112
Mesencephalon	.863*	.851*	.342*	-.179
Cerebellum	.874*	.877*	.358*	-.003

Medulla	.801*	.704*	-.136	.024

Note: * $p < .05$.

Post-hoc analyses for the cerebellum indicated that, as a function of brain size, its correlation significantly ($p < .05$) increases with the telencephalon ($r = .335$), with the diencephalon ($r = .381$), and with the mesencephalon ($r = .391$) when using body size residuals, while with raw size its correlation with other brain regions does not change as a function of brain size. When using phylogenetic contrasts, the relationship of residual cerebellum size with residual mesencephalon size is observed to increase as a function of residual brain size ($r = .240$), but no other significant effects were detected.

Similarly, post-hoc analyses for the mesencephalon indicated that, as a function of brain size, its correlation significantly ($p < .05$) increases with the telencephalon ($r = .318$), with the diencephalon ($r = .318$), and with the cerebellum ($r = .391$) when using body size residuals, while with raw size its correlation with other brain regions does not change as a function of brain size. When using phylogenetic contrasts, the relationship of absolute mesencephalon size with cerebellum size is observed to increase as a function of brain size ($r = .240$), but no other significant effects were detected.

3.2. Average Factor Loadings and Anatomical integration Coefficients for Telencephalic Structures

Examining the correlations among telencephalic structures, most regions are on average related importantly (see Table 3.3). This is observable both without and with phylogenetic controls, indicating that they are correlated across evolutionary time. However, no *changes* in the

factor loadings occur as a function of overall telencephalon size, as anatomical integration coefficients (reflecting the correlation of telencephalon size with factor loadings for a given brain region) for all regions exhibited low and non-significant magnitudes (Table 3.3). The omnibus test of integration (i.e., the overall relationship of telencephalon size with CPEs in general) yielded a weak and non-significant result of .111 and of .180 respectively ($p > .05$). The only possible effect was the bulbus olfactorius accessorius, which exhibited *weaker* relations to other regions in species with a larger telencephalon, although this was not replicated using phylogenetic controls. Evidence was found that regions that are on average less central to the common factor (i.e., having weaker average factor loadings) exhibited weaker or more negative anatomical integration coefficients ($r = .967$, $p < .05$; but $r = .434$, $p > .05$ after phylogenetic controls).

Table 3.3.

Average factor loadings and anatomical integration indices for regions of the primate telencephalon, without and with phylogenetic control

Brain region	**Factor loading**	**Evolutionary factor loading**	**Anatomical integration coefficient**	**Evolutionary anatomical integration coefficient**
Neocortex	.883*	.942*	.185	.172
Hippocampus	.955*	.926*	.102	.296
Striatum	.767*	.469*	.153	.145
Bulbus olfactorius	.604*	.666*	.065	-.258
Bulbus o.	-.135	.451*	-.332*	.113

accessorius				
Lobus piriformis	.961*	.955*	.170	.216
Septum	.951*	.966*	.183	.197
Schizocortex	.967*	.963*	.164	.201

Note: * $p < .05$.

Equivalent analyses to those presented in Table 3.3, but with all measures being residualized against body size, revealed a similar pattern. The volume for most, but not all, telencephalic structures exhibited strong and significant average factor loadings, and almost no changes in factor loading strength as a function of telencephalic volume were detected. The exceptions were the bulbus olfactorius, which is less correlated to other regions in species or larger telencephalon once phylogenetic controls are employed; and the lobus piriformis, which appears more correlated to other regions as a function of telencephalon size once phylogenetic controls are employed. The omnibus anatomical integration coefficient was weak and nonsignificant both for the raw and for the evolutionary approaches (.120 and .153 respectively, $p > .05$). Tentative evidence was found that regions that are on average less central to the common factor (i.e., having weaker average factor loadings) exhibited weaker or more negative anatomical integration coefficients, but this relationship was nonsignificant ($r = .401$, $p > .05$; $r = .600$, $p > .05$ after phylogenetic controls).

Table 3.4.

Average factor loadings and anatomical integration indices for regions of the primate telencephalon residualized against body size, without and with phylogenetic control

Brain region	**Factor loading**	**Evolutionary factor loading**	**Anatomical integration coefficient**	**Evolutionary anatomical integration coefficient**
Neocortex	.855*	.405*	.157	.134
Hippocampus	.806*	.755*	.175	.156
Striatum	.044	.146	-.046	.195
Bulbus olfactorius	.155	.473*	-.051	-.403*
Bulbus o. accessorius	.438*	.415*	-.001	-.065
Lobus piriformis	.864*	.846*	.150	.320*
Septum	.904*	.824*	.188	.250
Schizocortex	.880*	.858*	.175	.168

Note: * $p < .05$.

Post-hoc analyses for the lobus piriformis indicated that, as a function of brain size, its correlation significantly ($p < .05$) increases only with the hippocampus ($r = .299$) when using body size residuals, while with raw size its correlation with other brain regions does not change as a function of brain size. When using phylogenetic contrasts, the relationship of absolute lobus piriformis size with hippocampus size is observed to increase as a function of telencephalon size ($r = .368$) and to decrease with bulbus olfactorius size ($r = -.306$), and when using phylogenetic

contrasts along with body size residualization, telencephalon size positively predicts the relationship of lobus piriformis size with hippocampus size (r = .360), septum size (r = .385) and with schizocortex (r = .329), and negatively predicts the relationship of lobus piriformis size with bulbus olfactorius size (r = -.483).

The correlation of the bulbus olfactorius accessorius with all other regions was lower in species of larger telencephalon (ranging from r = -.341 to -.391, p < .05), except with the bulbus olfactorius, using the non-residualized, non-phylogenetic approach. No significant relations between telencephalon size and the magnitude of the association between that telencephalic region and others was detected when using either residualization against body size, phylogenetic contrasts, or both. Somewhat similarly, the correlation of the bulbus olfactorius with all other regions was lower in species of larger telencephalon (ranging from r = -.425 to -.496, p < .05) except with the bulbus olfactorius accessorius and the striatum, using data residualized against body size and phylogenetic contrasts. No such effects were detected in the other analytical approaches.

3.3. Post-hoc Analyses with on-Average Related Telencephalic Regions

It is immediately perceptible that there were low (and in some designs, non-significant) factor loadings for the striatum, bulbus olfactorius and accessorius in factor analyses of telencephalic regions. The decision to keep them in the analyses, however, stems from the aforementioned realization that a factor loading identified from traditional factor-analytic methods refer only to the *average* part-whole correlation between the volume of a neuroanatomical region and the volumes of the set of regions. As using CPEM has precisely the goal of examining whether factor loadings are *higher than average* in some primate lineages

(and thus perhaps moderate and significant for the striatum, bulbus olfactorius and accessorius in some taxa), excluding these regions from analyses altogether would defeat the purpose of the present study.

Nevertheless, in order to demonstrate that not excluding the three regions with low average factor loading does not importantly affect results, and for full disclosure, all analyses for the subregions of the telencephalon were repeated in full, but keeping these three regions out. Results are included in Appendix C, and were largely replicated in spite of these exclusions. The only exception was a slight increase (and crossing of the $p < .05$ threshold for significance) for the integration index for the hippocampus, and only in one of the four additional analyses presented in Appendix C. This suggests that, in species with a larger telencephalon, the volume of the hippocampus may be more strongly related to the volume of the other four regions kept in these analyses, but not with the striatum, bulbus olfactorius, and accessorius.

4. Discussion

Many studies have aimed to elucidate the relationships among brain regions in terms of how their sizes evolve in primates and other mammals. They involve tests of mosaic evolution hypothesis (i.e., regions changing in size through evolutionary time independently of others), and alternatively of concerted evolution. Few papers have attempted to examine differences among species in the degree of mosaicism or proportionality in the evolution of volumes of regions, and they have done so by examining clusters of species separately, losing statistical power severely. The present study aimed to employ a relatively novel statistical approach (especially within the field of phylogenetic comparative methods) to examine if species with a larger brain exhibit more concerted evolution among components of the brain. This also had the goal to test if such

an integration of volumes among brain regions could be the factor behind the demonstrated increases in integration among cognitive abilities in primate species of higher intelligence (Chapter 2).

Unlike the clear and consistent pattern of integration observed for cognitive abilities as a function of *G* in Chapter 2, no such pattern was identified for brain regions as a function of brain size, or for telencephalon subregions as a function of telencephalon size. Although a few positive results emerged (discussed below), the overall trend appears to be for neither stronger nor for weaker correlations among neuroanatomical components as a function of brain size or telencephalon size.

In spite of the considerable level of nesting (which introduces complexity) in the analyses presented above, their meaning and possible interpretations are straightforward: The general lack of effect of brain size upon correlations among sizes of brain regions, barring a few exceptions, means not that the sizes of regions vary independently of each other. The positive and strong factor loadings already demonstrate clearly that the regions are highly correlated. The lack of effect of brain size upon these correlations simply indicates that they do not become stronger or weaker as a function of brain size. In other words, the results of the factor analyses represent well the strength of relations among brain regions in terms of their size, irrespective of the brain size of the taxon in question. The logic also applies to relations among subregions of the telencephalon: the magnitude of their associations is well represented in the results of factor analyses, not varying considerably as a function of telencephalon size.

Smaers and Soligo (2013) have demonstrated that mosaic changes in neuroanatomy differentiate great apes (including humans) from other primates. Great apes also happen to comprise lineages that evolved larger brains. Contrary to hypotheses I and II outlined in this

study, this could suggest that the evolution of larger brains in primates involves more size independence among brain regions. In contrast, present results suggest neither stronger mosaic nor more concerted evolution in lineages that evolved bigger brains. In other words, sizes of brain regions are associated to each other across primates (suggesting a lack of mosaic evolution) in a constant strength, even after accounting for body size; that is, the strength of the associations among sizes of regions do not tend to become stronger or weaker as a function of brain size. Similarly, the correlations among subregions of the telencephalon are largely the same across taxa of different telencephalon sizes, suggesting neither a stronger mosaic nor a stronger concerted evolutionary pattern as a function of telencephalon size. Instead, regions exhibit a largely constant concerted pattern (as factor loadings are strong), with the exception of bulbus olfactorius, bulbus olfactorius accessorius, and the striatum. Thus, if clade specific variations in the importance of mosaic evolutions do exist as suggested by Smaers and Soligo (2013), brain size and telencephalon size do not appear to be the drivers behind that phenomenon.

4.1. Few Inconsistent Exceptions: Cerebellum, Mesencephalon, Lobus Piriformis, and Bulbus Olfactorius

Further analyses involving a larger number of species would be ideal to test whether the putative integration effects identified for the cerebellum, mesencephalon, lobus piriformis, and hippocampus are robust, and whether the disintegration of bulbus olfactorius and accessorius from other structures are robust as well. Considering their lack of consistency across methods (i.e., residualized and non-residualized, controlled for phylogenetic non-independence among species and not controlling for it), and considering the lack of significant omnibus integration effects throughout analyses presented in this study, any interpretations made for these regions

needs to be considered highly tentative and necessitates further examination. A review of the possible reasons for these effects identified is made below based on the primate neuroanatomy literature, although present results are highly tentative at best.

Cerebellum and mesencephalon, when residualized against body size, appear to exhibit stronger correlations to each other and to other brain regions as a function of brain size. This was not replicated once phylogenetic contrasts were used. These results possibly suggest that (a) larger brains and (b) stronger correlations of the cerebellum and mesencephalon to each other and to other regions, did not coevolve *in lockstep*, but an overall association appears to exist that perhaps evolved out of tempo.

The current analyses reflect that, as a function of brain size, the cerebellum specifically is tentatively more strongly correlated with the telencephalon, the diencephalon, and the mesencephalon. These possible results make sense considering the fact the cerebellum is involved in multiple functions besides motor performance (e.g., coordination, planning, and motor adjustments; Vanderah & Gould, 2015), including cognitive abilities, and the reception of visual and auditory information (Petacchi et al., 2005). These processes require a constant exchange of information, evidenced by the multiple tracts reaching this area. For example, the cerebral cortex sends and receives information to the cerebellar nuclei (Vanderah & Gould, 2015), and the diencephalon (more specifically the ventral lateral and the ventral anterior nuclei of the thalamus) is also linked with the cerebellum (Hoshi et al., 2005; Middleton & Strick, 2001), acting as intermediaries in the transmission of motor information between the cerebellum and the cortex in primates.

Concerning the apparent increasing cerebellar-mesencephalic association in larger brains, the strongest effect observed in the present analyses, it is possible that some subregions of the

mesencephalon exhibiting direct paths to the cerebellum may be behind this positive and significant correlation identified. The tegmentum, for example, is connected with the cerebellum via the superior cerebellar peduncles. It also contains the red nucleus, a neural hub involved in motor coordination and receiving cerebellar projections via the interposed nucleus and the inferior olive (Vanderah & Gould, 2015). Even though the current comparative literature only provides data for some neuroanatomical regions, future studies should determine whether the volume for areas such as the cerebellar peduncles and the tegmentum are more correlated to evolutionary changes in the cerebellum as a function of brain size.

The hippocampus exhibited significantly higher correlation to the sizes of other regions of the telencephalon in species with a larger telencephalon, but the effect is weak and only present once the striatum, bulbus olfactorius and accessorius are eliminated from the analyses. Interpretations should be made with caution. It is possible, however, that this tentative effect reflects the multiple and important connectivity that the hippocampus has with other brain regions. The hippocampus has afferent and efferent fibers to multiple areas of the neocortex and to subcortical forebrain structures through the entorhineal cortex (Cammarota et al., 2005; Moser & Moser, 1998). Connections of the hippocampus with other regions are complex, as the posterior and anterior hippocampus areas exhibit different connections and thus help control many cognitive and affective processes (Rocca et al., 2018).

On the other hand, the bulbus olfactorius and the bulbus olfactorius accessorius appeared to exhibit markedly lower correlations to other structures of the telencephalon in species with a larger telencephalon, though these results were method-specific. Interestingly however, the only structures to which they remain as strongly correlated in species of larger telencephalon are each other, and in one particular approach, the striatum. This reinforces the notion that the striatum is

somewhat connected to the bulbus olfactorius in function (Newman & Winans, 1980; e.g., with the striatum using information obtained through olfaction to guide behavior; Setlow et al., 2003). Note that these three telencephalic regions were the three for which comparatively low factor loadings in the unit-weighted factor estimation were observed overall, compounding the notion that they covary to a certain degree independently of other structures of the telencephalon in terms of their size.

The lobus piriformis appeared to exhibit higher correlation to the sizes of other regions of the telencephalon in species with a larger telencephalon, but this is most likely not driven by cognitive evolution, given lack of indications of a role for the lobus piriformis in cognitive skills. Moreover, this stronger correlation with other regions in terms of volume is paradoxical to the aforementioned weaker correlation of other regions responsible for processing olfaction (i.e., the bulbus olfactorius and bulbus olfactorius accessorius). Unlike the relative independence of these latter regions from others in the telencephalon, it is possible that information from the lobus piriformis is being recruited for cognitive and affective processing in other regions of the telencephalon.

4.2. Future Further Tests of Theory

Future work from this research program will involve examinations of possible integration among sizes of brain regions as a function of *G*, however such work is in initial steps as larger, more overlapping datasets are needed, and thus is in the process of preliminary collation and analyses. Expansions to datasets of neuroanatomical volume measures would also increase reliability of any future estimates calculated – only in 2018 an expansion to Stephan and colleagues' (1981) dataset was produced, and contributions to this field are unfortunately sparse.

Preliminary analyses of the relationship of *G* with the strength of correlations about volumes of brain regions indicate a lack of effect just as the effect of brain size upon correlations as presented in the present study. However parameter estimates in this preliminary study of the effects of *G* upon correlations among brain regions exhibit high standard errors considering (a) the small overlapping sample of *G* and neuroanatomical measures, and (b) the several controls that need be employed (for body size allometry and phylogenetic non-independence).

4.3. Other Possible Forms of Integration

A stronger correlation among performance scores on different problem-solving skills in primate species of higher general intelligence can potentially be explained by multiple different neuropsychological hypotheses. While the present study indicates that increased correlation among brain regions in species of bigger brains does not appear a likely avenue, other putative explanations exist. Firstly, it is possible that the increased integration observed among cognitive abilities reflects more effective communication among diverse brain regions. This would involve increases in projections across regions, and thus in white matter, but not necessarily concerted changes in the volume of those regions. Functional connectivity is a known mechanism for the coordination of activity among different neural assemblies in order to achieve a complex cognitive task (Fingelkurts et al., 2005; Sporns et al., 2000).

Note however that this connectivity hypothesis for explaining the integration of cognitive skills in species of high general intelligence is not fully independent from the hypothesis of neuroanatomical correlation tested in this study, because regions that communicate more intensely through projections also tend to covary in their volumes (Park & Friston, 2013; Whiting & Barton, 2003). Furthermore, long-range connectivity has been noted to, on average,

decrease in species of larger primate cerebral cortices, and thus connections in such brains tend to be concentrated among contiguous, not distant areas (Herculano-Houzel et al., 2010). Such small-world networks potentially characterized by so-called "proximity-dependent modularity" due to short-association fibers in large primate brains (Gómez-Robles et al., 2014) counter the likelihood that it is increased *long-range* connectivity that is behind the observed integration among cognitive abilities in species of high general intelligence (demonstrated in Chapter 2).

On the other hand, interconnectivity via short association fibers communicating adjacent regions has been proposed to be behind the advanced cognitive capacity of apes and especially humans (Schenker et al., 2005). Furthermore, even for longer-distance projections, increased myelination may, to some degree, compensate for reduced long-range connectivity (Schenker et al., 2005). As discussed in Chapter 1 however, these alternative explanations to the integration of cognitive abilities, involving connectivity and myelination, remain to be rigorously studied through the comparative approach, and are still yet to be related empirically both to general intelligence and the degree of integration among cognitive abilities.

4.4. Limitations and Future Directions

It must be noted that few specimens per species were included in the original data collection processes, given the difficulty and costs of detailed volumetric measures of structures of fresh and intact brains for, in many cases, primate species of comparatively rare and difficult access. This lack of representativeness may lead to biased estimates of average volume for many brain regions.

Potential sources of bias in analyses and interpretations of their meaning involve lack of information on neuron density in most brain regions. Although a few studies exist, they involve

few species and few brain areas, not being fit for generalizable conclusions, as different regions exhibit different changes in neuron density as a function of brain size (Barton, 2012). As such, even though correlations between volumes of brain regions exist and may change as a function of brain size as tentatively demonstrated for a few regions in the present study, this does not mean that necessarily their numbers of neurons are proportionally increasing as well. Likewise, the correlations between brain regions in terms of their volumes are difficult to interpret with precision given limited information on how their interconnectivity varies. Future studies that address these limitations or provide more data for primate species, indirectly permitting further better-informed neuroanatomical analyses and interpretations, would be of great value. Further data that also increase the overlap between datasets on intelligence and neuroanatomical region volumes would be likewise essential.

Integrative and Concluding remarks

1. The Evolutionary Trajectories of *G* and of Neuroanatomical Volume Measures

In cross-species analyses with primates, comparisons of general intelligence to potential substrates to examine which neurobiological structures account for it, had before the present studies relied on the correlative approach. As illustrated in the Introduction Section, even variables where species exhibit radically different absolute differences may exhibit a perfect correlation, because the *r* coefficient simply indicates whether the *proportions* among values within the two variables are comparable. As such, if forelimb size increases .1 cm per generation and hindlimb size increases 10 cm per generation consistently in a hypothetical lineage, the evolution of forelimb size and hindlimb size are *perfectly* correlated even though hindlimb size is under selection for more much more radical changes. Therefore, examining the evolutionary rates of *G* and of the most commonly used neuroanatomical volume measures (NVMs) was imperative to test this non-explored degree of comparability between variables. In other words, correlation is not sufficient as an indicator of comparability in terms of evolutionary trajectories.

Considering that not all measures examined in the preceding chapters were in the same metric (i.e., *G* is not measured in cubic centimeters as NVMs), their evolutionary rates were compared in terms of standardized (*z*-score) changes per million years. Limited comparability was found, with *G* evolving much faster than most NVMs. The exception of the cerebellum, and to a lesser degree neocortex size, was identified once they were residualized against body size, but even those regions evolved at most half as fast as *G*. Moreover, *G* was identified to have accelerated its evolutionary rate considerably more than even cerebellar and neocortical size, and considerably more than other NVMs.

Furthermore, a point can be made that, even when correlations between hypothetical variables are moderate to high (for instance, around .5), it is possible that their considerable unique variance (in this hypothetical case, around 75%) means that they have been exposed to different selection regimes, of partly in different directions, or that one variable exhibits evolutionary changes in a consistent direction while the other more strongly fluctuates in a somewhat random walk. In Chapter 1, variables were compared in the fit of evolutionary regime models. Absolute (i.e., non-residualized) NVMs exhibited significant fit to the early burst model, whereby niche filling occurs early in the phylogeny and subsequently few changes occur in the daughter lineages. In contrast, not only did *G* exhibit positive acceleration (increasing rather than decreasing rates of changes), but it also exhibited a strong fit to the Ornstein-Uhlenbeck model, whereby trait changes across evolutionary time in most lineages occur toward a direction and an optimal level. While it is clear that changes in *G* were not towards increases in all lineages at all times (with some exhibiting reductions in *G*; Reader et al., 2011), overall this trend toward a higher optimal level for *G* identified in Chapter 1 was much stronger than the trend observed for any NVM. Taken together, findings from Chapter 1 further compound concerns on reliance upon the simple correlative approach, and suggest that no NVM alone is comparable to *G* in terms of evolutionary history.

Chapter 2 showed that increasing *G* coevolved with increasingly strong associations (operationalized as factor loadings) among cognitive abilities. The same was not observed for brain regions in Chapter 3. Namely, at both levels of analyses conducted (the entire brain, and the telencephalon) regions were not more or less strongly correlated as a function of overall size. Some exceptions emerged, but not in a strong or consistent fashion across the methods used. Thus, even though *G* and NVMs exhibit correlations (Deaner et al., 2000, 2007; Dunbar, 1992;

Shultz & Dunbar, 2010), the evolution of their internal structure is not highly comparable. In other words, while the strength of the positive manifold for *G* has varied predictably and in line with *G* levels themselves, the same is not true for the positive manifold of brain regions or telencephalic regions. The strengthening of associations among cognitive abilities may be due primarily to other neurobiological factors to be explored further, such as possibly inter-region connectivity.

Just as the encephalization quotient contributed, in the history of the academic study of the neurological bases of intelligence, to a complex understanding of species differences in the evolution of the brain and overall cognitive processing power, but eventually being deemed suboptimal in light of new findings and giving way to alternative measures (Falk & Gibson, 2001), similarly NVMs, either in their absolute metric or controlled for the covariation with body size, seem to show signs of strong limitation. Other measures can be argued to be necessary to further understand not only correlation but shared evolutionary processes and causes with intelligence.

2. Substrates for General Intelligence: Beyond the Traditional NVMs

The exceptions to the observation that NVMs are not comparable to *G* in evolutionary patterns were residual cerebellar volume and, to a lesser degree, residual neocortical volume. These two measures showed moderate similarity to *G* in their evolutionary history, in Chapter 1. Previous evidence discussed at length in Chapter 1 and Chapter 3 indicate that cerebellar size is indeed greatly underrepresented in studies of substrates of intelligence, as it has coevolved with cognitive abilities, increased in size especially in apes, and has not exhibited decreases in neuron density in species with larger cerebellum. While it is not known precisely how the cerebellum

permits or enhances cognitive abilities, it is likely that (a) it acts as an enhancer or augmenter of the activity of cerebral regions involved in cognition (Leiner et al., 1989; Snider, 1950), or that (b) it facilitates motor coordination (Doyon et al., 2003; Salman, 2002), permitting technical skills such as tool use, extractive foraging, and thus manipulation of environments and ultimately niche construction. Interestingly, in Chapter 3, among the few regions that exhibited some signs of stronger correlations with other brain structures as a function of brain size was the cerebellum. Although this particular result must be interpreted with caution given the lack of consistency of integration effects in Chapter 3, it would be in line with recent evidence suggesting that the cerebellum is intensely connected to other regions and thus interacts with their functions (Vanderah & Gould, 2015). Connections with the neocortex are especially notable, perhaps compounding the findings, from Chapter 1, that these two regions to a moderate extent have had similar trajectories to those observed for *G*. It is possible that these (and perhaps other NVMs) *together* account for a larger portion of the evolutionary trajectory of *G*, even though separately none has exhibited strong similarity to *G* in Chapter 1. This notion is counter to the tendency, in primatology, to attempt to find which *one* NVM accounts for intelligence.

Alternatively, several non-volumetric measures have been proposed as permitting complex cognition (Barton, 2012; Dicke & Roth, 2016; Hofman, 2001; Roth & Dicke, 2012). Examples are neuron density, myelination, gyrification, and their case is reviewed at length in Chapter 1. Little data exists for robust examinations among primates however, and it is beyond the scope of the present work to forward speculations about their importance without careful analysis. It must at least be noted that, while it is improbable that any of them individually fully accounts for the evolutionary history of intelligence, a shift of further attention towards them may help explain primate *G*.

Future initiatives to examine the phylogenetic history of *genes* implicated on *G*, NVMs, and other neurobiological measures can also be highly informative (Enard, 2014; Montgomery & Mundy, 2012). Extensions of such studies to other neurological indicators could do a great service to understanding the evolutionary trajectories of potential substrates of intelligence. Like studies that identified “generalist genes” behind multiple cognitive capacities in humans (and thus possibly behind general intelligence; Kovas & Plomin, 2006), the degree of pleiotropy in genes behind NVMs and other neurobiological traits would further illuminate the degree of comparability between general intelligence and neurobiological measures.

It is possible that the future study of these or other aspects of the brain may account for the high, accelerating, and largely directional evolutionary rates of *G*, but it is also possible that no aspect of the brain has seen comparable evolutionary rates. It may be that, instead, intelligence has accelerated its evolutionary rate in many primate lineages through a runaway process that does not require commensurate increases in cognitive processing; this process has been termed "ratcheting" (Tennie et al., 2009). While most discussion of the ratcheting effect has focused on making a binary contrast between humans and other primates (Last, 2014; Tennie et al., 2009; Tomasello, 2011) in the rate of complexification of cultural developments derived from cognitive skills, it is also possible to think of the ratcheting phenomenon as having gradations (being more extreme in humans, but perhaps present in varying degrees in other species). Humans’ cognitive feats have exponentially increased in a runaway complexification that seems to greatly surpass rates of change in brain size or other neuroanatomical measures. This is possible because we do not need to reinvent the wheel in each generation – previous innovations, insights, and solutions to problems are accumulated, socially transmitted, built upon, and combined for exponentially advanced developments that are not made by single

individuals, but in complex collaboration. As such, to an observer who simply took a modern cross section of the state of human problem-solving, it could seem like humans possess incredibly advanced cognitive skills that would necessitate a much more advanced biological substrate, if we were to achieve in a single generation everything that we are capable of achieving because of cumulative and transmissible knowledge.

Although authors discussing ratcheting in humans hypothesize that humans are the only species to surpass the threshold necessary for this runaway process (Laland, 2018), it may exist at a slower rhythm in other primates, especially apes, although to a much less extreme extent. It is possible that in non-human primate species innovations, coupled with social learning (in fact two of the measures used in Chapters 1 and 2 as indicators of *G*), permit this emergent property of culture whereby more complex problem-solving is permitted because of previously existing cognitive resources in the cultural pool. Such a ratcheting effect would not necessarily require a commensurate and in-tandem increase in brain volume or neural characteristics enhancing processing power. Accelerating innovation rate may be possible in species of sufficiently high *G* simply because previous innovations exist and have been socially transmitted. As such, the increasingly evolvability of *G* reported in this study may partly reflect runaway increasing possibilities of use and combination of previous solutions transmitted socially.

3. The Validity of *G*: Signs of Adaptive Design

Blomberg and colleagues (2003) seminal study identified that behavioral measures exhibit lower values in tests of phylogenetic signal (i.e., phylogenetic inertia or conservatism), and posited that perhaps error in measurement in behavioral traits could be behind this observation. Anatomical and biometric life history measures, in contrast, were presumed to be

more reliably assessed and also exhibited more conservatism. If this were true in the case of *G* in Chapters 1 and 2, whose observable indicators were also behavioral, *G*'s evolutionary history would show signs of random variation, thus exhibiting (1) no consistent acceleration or deceleration in evolutionary rates, (2) no direction towards which most of changes occur. Moreover, if the patterns identified for *G* reflected measurement error, this would have been identified especially by the λ and *K* estimates, which would be precipitously lower for *G* than for other variables. In fact estimates for some NVMs were even lower than those for *G*. The results presented in Chapters 1 and 2 demonstrate that the low evolutionary conservatism on *G* in primates is unlikely to be the product of measurement error: Very high acceleration has occurred in its rate of change on average, and a clear pattern of directional selection towards a higher optimum emerged. Moreover, selection regime differences using this dataset appear strongest when contrasting apes and other primate species (Fernandes, Peñaherrera-Aguirre, Woodley of Menie, & Figueredo, in preparation), a finding converging with previous discussions of primate intelligence using other data (Deaner et al., 2006; Parker, 1978; Reader et al., 2011; MacLean et al., 2014; Seed et al., 2009). The evolutionary history for *G* was clearly not one of random variation and the dataset employed in the present studies exhibit signs of reliability and convergence with other sources.

The low conservatism identified for *G*, therefore, is much more likely due to quick differentiation from ancestor lineages, along with the observation that higher general intelligence evolved in multiple primates lineages (Reader et al., 2011). As multiple lineages exhibited convergent evolution, closely-related species are not extremely more similar to each other than more distantly-related ones in their level of *G*. This is, by definition, the core of low evolutionary conservatism (Felsenstein, 1985).

These indications that the evolutionary history of *G* exhibited directional, accelerating evolution in multiple lineages are signs of adaptive design for general intelligence. A hypothesis that *G* is an epiphenomenon or a by-product of other cognitive processes without function would lack parsimony (for a review and critique of the by-product hypothesis, see Burkart et al., 2017). This especially appears to be the case when it is realized that:

(1) Cognitive skill measures on which *G* loads more strongly exhibit faster evolutionary rates and less conservatism (Fernandes et al., 2014), thus the locus of selection appears to be *G* itself, not the unique variance in diverse cognitive skills; and

(2) As demonstrated in Chapter 2, there has been coevolution between level of *G* and the strength of its manifold, suggesting that as selective pressure pushed for more *general* intelligence, the more *generalized* (i.e., incorporating, or drawing from, multiple specific abilities) it became.

Considering these multiple lines of study altogether, the longstanding concern that a common factor among cognitive performance measures may simply reflect an artifact of measurement and does not measure a true latent construct (and thus has no functional, adaptive design; for a review, see Jensen, 1987) also loses credibility considering the multiple signs of non-random, non-neutral evolutionary processes behind *G*. Such concerns about measurement validity have also been placated previously in reports of high comparability between laboratory and ethological measures of general intelligence (Burkart et al., 2017; Reader et al., 2011) and in agreement of these approaches with expert-based rankings of primate species (Reader et al., 2011).

4. Implications to Evolutionary Psychology

Insights gathered from the present set of studies and from the blossoming recent literatures on general intelligence altogether suggest that:

(1) Rather than a single or newly developed ability, general intelligence appears to recruit from existing abilities, based on findings from Chapter 2: Species of high general intelligence integrate several skills that are individually and more independently present in taxa of lower general intelligence.

(2) Rather than simply complementary but independent modules exhibiting impenetrability and autonomy in a Fodorian sense, it appears that cognitive skills interact and that there is considerable transfer among them (de Waal, 2016); that is, solutions applied to one specific cognitive problem are translated to solutions to different cognitive problems (e.g., from abstract reasoning to social reasoning; Girotto & Tentori, 2008) or to different learning contexts even if only partially.

(3) Rather than dealing with an adaptive problem of limited sphere, it appears that general intelligence in primates involves problem-solving in several aspects of foraging, solving technical challenges, and navigating the social world. Clearly, the scope of adaptive problems dealt with by general intelligence is *finite*, otherwise each and every psychological measure would be part of its manifold (including personality traits, life history characteristics, perception-related traits, sexuality tendencies, etc.). In that sense, general intelligence has a *limited* (in the sense of finite), but *broad* domain. However, this domain is not specific to one adaptive problem; rather it is arguably general, including situations that present novelty in diverse ways.

(4) Finally, rather than evolving specifically in ancestral environments in human evolutionary history, it is clear that increasing (a) levels of general intelligence (Deaner et al.,

2006; Fernandes et al., 2014; Reader et al., 2011) and (b) integration among cognitive abilities (see Chapter 2) evolved in multiple lineages in the primate phylogeny, with accelerated evolutionary rates (see Chapter 1) not being limited to the human lineage.

The present findings contrast with the tendency, in evolutionary psychology, to frame adaptations as arising in response to domain-specific problems of narrow scope. For instance, Kanazawa (2004, p. 512) proposed that general intelligence is a true adaptation but that it "evolved as a domain-specific adaptation for the originally limited sphere of evolutionary novelty in the ancestral [human] environment." Instead, the present findings and the broader comparative literature they are inserted in suggest that general intelligence evolved as a set of adaptive, increasingly integrated cognitive systems for the complex and broad sphere of evolutionary novelty in many primate lineages. Preliminary findings that habitat generalism and diversity is associated with both *G* and with stronger relations among cognitive abilities in primates (Peñaherrera-Aguirre & Fernandes, in preparation) strongly compound the notion that novel challenges were not an uncommon problem for many primate lineages. Moreover, several cognitive abilities included in the *G* factor in the present studies appear associated with success rate in artificial introduction to novel environments and with invasion rates, suggesting that they do facilitate navigation of novel niches (as reviewed in Chapter 2), which, by definition, present evolutionary novelty.

The common claim, in evolutionary psychology, that a class of situations must be narrow and specific (rather than general and novel or varying) to exert consistent selection pressure is insufficiently justified (Kaufman et al., 2011). Any environmental regularity may exert selection pressure, as long as it poses a challenge or opportunity. The rise of an adaptation simply reflects the likelihood that genetic variation might lead to variation in the ability to address the challenge

or take advantage of the opportunity. This logic applies just as well to the class of *novel* problems: Regardless of the fact that they are apparently distinct, selection can exist for increased general ability to identify the structure of the novel problem, to find novel solutions or abstract and transfer from solutions applied to other, previous problems (de Waal, 2016).

It is important to stress that the evidence for the existence of *G* and for its interaction among cognitive skills does not necessarily negate plausibility for the massive modularity account from evolutionary psychology. In contrast to a Fodorian view of modules, massive modularists have repeatedly proposed that encapsulation and impenetrability among modules are not necessary (Tooby et al., 2005). In discussing interactions among modules, Barrett and Kurzban (2006, p. 633) also exemplify that "for systems involved in inference, reasoning, judgment, and decision making [...] —and indeed, for most systems other than very early perceptual processing—there is every reason to expect both interactivity and the integration of information from multiple sources." Furthermore, it is argued that modular systems permit components to be combined in novel ways, not precluding flexibility to tackle novel challenges or opportunities. The present studies exhibit evidence that performance on different problems became more strongly correlated across primate evolutionary history, potentially suggesting combination and interaction among already existing (and apparently more independent) cognitive abilities. It remains to be examined, as a future step, how such combination takes place – possibly through stronger connections among neural systems, or through an integrator or centralizer that organizes and mediates their communication. Possibilities are further discussed below.

5. The Nature of General Intelligence

It must also be noted that, as informative of the evolution of *G* as this study might be, considering that the present work included only species mean scores, and no within-species variance, it is impossible to make assertive inferences about the evolution of *g* (the individual differences level general factor of intelligence) for any particular species based on these analyses and its relation to neuroanatomical regions. It is unwise to forward assertive conclusions about whether the nature of *G* is commensurable to the nature of *g*, as the neural characteristics that differentiate species in cognitive performance may or may not be the same as those that differentiate individuals (Arden & Zietch, 2017). To assert that *G* and *g* are reflective of the same phenomenon, further research is needed to examine whether their nomological nets overlap (Burkart et al., 2017) and if differences among individuals of different species also lead to a general factor when a mixed (i.e. both species and individual level data) approach is used (as suggested by preliminary evidence; Woodley of Menie et al., 2017).

As acknowledged by Burkart et al. (2017), the methods presently available do not permit confidently asserting *how* general intelligence exists in the brain. As mentioned above, it may be that general intelligence exists as strongly connections among neural systems, or perhaps an integrator exists that recruits from several systems. The findings from the present work and from recent literatures, summarized above, suggest that some information is available about how general intelligence is organized in terms of its nature. However, much about it is still unclear. It is even hypothetically possible that rather than being a latent trait reflecting an integrated system, general intelligence *assists* in the performance of multiple specific cognitive abilities that operate on multiple domains. Even if general intelligence arises as a result of process overlap because cognitive performance on various domains all tap the same domain-general executive process

(e.g., working memory), there is still covariation among them that is due to a cognitive process that assists in problem-solving in many types of adaptive problems. How exactly the positive intercorrelations arise (i.e., the nature of their relations) does not matter for claims that (a) they are correlated abilities, and that (b) general intelligence was a target of selection, whatever its nature (Burkart et al., 2017). Of course, understanding its nature is of interest, but is not a necessary step for examinations of the evolutionary history of its aggregate level and of the strength of its manifold.

A comparison of general intelligence to another abstract concept can help illustrative this point. Consider the case of movement speed. No single muscle or hormone is solely responsible for it just like trying to pinpoint a specific region in the brain responsible for general intelligence is an illusive endeavor. Multiple tissues need to be engaged for increased movement speed, thus speed derives from a system. No single gene is directly responsible for speed either, they simply lead to synthesis of molecules that will ultimate carry on various tasks in many tissues necessary for speed, among other things. However, debating whether movement speed is (1) a latent trait among the actions of tissues in the system, or (2) an emergent property of their actions, or (3) any other alternative conceptualization, does not invalidate the fact that movement speed exists as a phenomenon of interest and as a target of selection. Individuals and species can be compared on it. The important focus, at least at first, should arguably be (1) that different approaches to assessing it and its outcomes converge into the same similar patterns of individual and species differences, thus exhibiting validity, (2) establishing whether movement speed is an important biological capacity that affects genetic replication success and has been the target of selection, and (3) examining which biological structures best function as substrates for movement speed. In this analogy, the study of G has so far contained studies that help these three questions, with

early contributions addressing (1) (Burkart et al., 2017; Deaner et al., 2006; Fernandes et al., 2014; Reader et al., 2011), and the present work contributes especially for (2) and (3).

To that effect, in the case of general intelligence, steps taken in the present work are important to further clarify the nature of G. Not only is G replicable in primates (Burkart et al., 2017; Deaner et al., 2006; Fernandes et al., 2014; Reader et al., 2011), but now evidence exists that it has been a main and largely consistent target of selection (Chapter 1), that it has coevolved with stronger associations among its components (Chapter 2), and that explaining its evolutionary history necessitates more than the commonly used neuroanatomical volumetric approach (Chapters 1 and 3). Still, much further work is necessary in years and perhaps decades to come to further elucidate the nature and substrates of G and of within-species g, in spite of the growing number of pieces assembled in this puzzle.

Appendix A: Operationalizations of Ethological Measures of *G*

This appendix includes conceptual definitions, operationalizations, existing taxonomies of behavior, and examples of behavior for each of the five broad categories of problem-solving used as measures of *G* in Chapters 1 and 2.

1. Extractive Foraging

Extractive foraging refers to the manual acquisition of foods from hidden locations or protective matrices, from which they are therefore not obviously or clearly retrievable (King, 1986). Food items foraged extractively by primates include nuts, roots, tubers, shellfish, adult and larval invertebrates that are embedded within protective exoskeletons or that reside in mounds, hives, thorns, crevices, branches, under bark, or furrowed in leaves (Gibson, 1986; Melin et al., 2014). Invertebrates provide a valuable source of protein and fat, which are essential and often limiting resources to primates (Melin et al., 2014). In a few species, extractive foraging may also include eggs, bone marrow, and brains (Gibson, 1986).

Scholarly attention to this trait has increased, both in ethological and laboratory works, following the proposal of the extractive foraging hypothesis of cognitive evolution (Parker & Gibson, 1977). This hypothesis proposes that species that detect and extract hidden food items, though lacking specialized anatomy for that purpose, should have high sensorimotor coordination and high cognitive ability to compensate, in order to exhibit successful extractive foraging (Parker & Gibson 1977; Byrne 1995). Note, however, that a few species do possess specialized anatomy for extractively foraging one type of food item, such as the aye-aye's middle finger for tapping along branches and rotting logs, its fourth, long, finger for extracting

grubs out of holes on the brances, and its ears specialized for ecolocation of signals from the tapping.

Detailed ethological observation in gray snub-nosed monkeys (Xiang, Liang, Nie, Li, 2013) has permitted a taxonomy of extractive foraging behavior. Four categories of extractive foraging were identified in this species, a model yet to be replicated within other species and at the comparative level:

(i) Removing the casing of a plant (e.g., peeling bamboo);

(ii) Excavation (i.e., digging or surface scratching);

(iii) Prying or picking (e.g., removing bark to reach objects underneath it);

(iv) Searching under obstacles (e.g. overturning rocks or probing under them).

Published reports of extractive foraging also predict performance in response to novel extractive foraging tasks under experimental conditions in primates (Day et al., 2003), supporting the use of literature counts, as the database used here collated by Reader and colleagues (2011), to estimate species differences in extractive foraging.

Note that, for some acts of extractive foraging in certain species, tools are used. However, for the collation conducted by Reader and colleagues, a decision was made in such cases as to whether (1) each such specific report made in the literature was characterized primarily as tool use or primarily as extractive foraging (e.g., in case tool use was secondary, rudimentary, or unsuccessful), or if (2) such a distinction could not be made, and thus the report was not considered for inclusion in the database. This was intended to avoid inflation of the *G* positive manifold.

Categorization of species into the classes extractive forager and non-extractive forager was common prior to Reader and colleagues (2011) non-binary database of registering counts of extractive foraging reports per species. Still, the binary categorization, Reader and colleagues (2011) more fine-grained organization of data, and captivity data on researcher-developed extractive foraging tasks converge highly, corroborating the validity of the measure employed in the present study (Day et al., 2003).

Judgment of whether a given behavior pattern qualified in one of the five categories of behavior was made the expert author of each original article (Reader et al., 2011). For instance, behaviors originally described as “novel”, “innovative”, “creative”, “never seen before”, among close synonyms, were included in the “innovation” category. Similarly, behavior descriptions containing keywords that connoted that an individual learned about resources, the environment, or others from conspecifics (rather than by themselves) were included in the “social learning” category. As previously argued by Lefebvre and colleagues (1997) and Reader and colleagues (2011; Reader, 2003; Reader & Laland, 2001), this approach avoids subjective bias on the part of the researcher responsible for collating data for the database. It must be emphasized that the behavior reports were not derived from uninformed casual observations, but rather came from experienced, established scientists extensively familiar with their subjects (Reader, 2003). Furthermore, observers themselves have more information about the context in which behaviors occurred than Reader, Hager, and Laland had for the collation of the database, and the context matters for the categorization of the behavior: for instance, the act of placing a potato inside a small body of water (as observed in Japanese macaques; Kawamura, 1959) may be indicative of innovation, of tool use, or of deception, depending on the context in which it happens, and the particular purpose it serves in that specific behavior instance. If it is a novel occurrence, it may

be classified as innovation; if instead the behavior is common in the species (as it is today in Japanese macaques) but the behavior in the case of this particular instance served to conceal the food item from conspecifics, then it may be classified as deception; and so on.

Behavior descriptions and their categorization were admittedly sometimes broad enough to could satisfy more than one of the five broad categories. For instance, it was not uncommon to find descriptions of social tool use, innovative tool use, innovative deception, among other combinations. To avoid conflation and thus overestimation of the strength of the positive manifold, such cases were not counted as duplicates, but rather only counted once if one aspect predominated (e.g., either tool use or innovation in the case of innovative tool use), or deleted altogether from the database if no clarity existed (Reader et al., 2011).

2. Tool Use

Perhaps the most studied of the five categories of problem-solving, tool use is also perhaps the one with the most debated and inconsistent definition, up until recent decades. Perhaps most gravely, this initial variability in how to understand whether a behavior constitutes tool use or not certainly affected ethological reports, with comparatively low inter-rater reliability for certain behaviors, although for the vast majority there was high agreement (Shumaker et al., 2011).

Examples of variations in tool use definitions, as detailed extensively by Shumaker and colleagues (2011), include the notions that:

a) tool use involves the use of an external object as functional extension of mouth, break, hand or claw, in the attainment of a goal;

b) rather, the external object must be held;

c) the target of tool use behavior must be another object (i.e., the goal is reach it, or modify its form, position or condition) with the intention of improving its efficiency;

d) rather, the target of tool use can also be self, a conspecific, or another living being;

e) any useful definition must be narrower, to avoid broad definitions whereby passive objects, such as a tree, would be considered a tool in the case of an individual climbing it to reach a fruit; or such as the ground, in the case of an individual dropping an egg on the ground to break it.

The most widely used definition of tool use, however, states that it is "the external employment of an unattached or manipulable attached environmental object to alter more efficiently the form, position or condition of another object, another organism, or the user itself, when the user holds and directly manipulates the tool during or prior to use and is responsible for the proper and effective orientation of the tool" (Shumaker et al., 2011; for close definitions but with small variants in wording, see Beck, 1980; Amanat & Horton, 2008). Examples of tool use in primates conforming to this definition include the use of sticks to forage for invertebrates (van Lawick-Goodall, 1968); the use of stone hammers to crack open nuts, folded leaves for drinking water, poles as ladders, and arguably even branches and stones as weapons (Silva & Silva, 2016).

Having been a focus of interest for many primatologists in the field for decades, forms of tool use are commonly included in detail in ethograms, and detailed descriptions are available in the literature. Some examples illustrate the high level of care in the reports:

The behaviour consisted of three main steps: (i) the monkey approached the nest, reaching for and then tapping (firmly and quickly, as when capuchins tap hard surfaces)

> *the nest exterior immediately in front of him using both palms (when in front of the nest, the body in a squatting position; semi-prehensile tail used to anchor the body on a stable horizontal branch (10–15 cm in diameter)); (ii) the monkey tore off a branchlet (hereafter, stick) approximately 20–30 cm long from the tree where the nest was located (to tear off the stick, the capuchin may temporarily adopt another body position), perforated the nest with it (approximately in the middle) and then inserted it into the nest (5–10 cm deep).* (Souto et al. 2011, pp. 533-534).

Other highly detailed step-by-step descriptions include how orangutans use stick tools to extract honey from bees' nests (van Schaik et al., 1996); how they also use sticks to pry open and remove the hairs of Neesia fruits (van Schaik et al., 2003); how capuchin monkeys use hammers and anvils to crack open nuts and seeds (Aguiar et al. 2014; Fragaszy et al., 2004); and how long-tailed macaques use stones in two alternative but highly organized ways to break oysters (Gumert et al., 2009).

Other times, especially with uncommon tool use behavior that is not part of an organized ethogram, descriptions are highly anecdotal. Admittedly, such observations might be accidental tool use or even overinterpretation on the part of the primatologist, whereas in reality the behavior was a happenstance. Examples abound, especially for complex behavior: A Western lowland gorilla was observed using a handheld detached tree branch seemingly as a walking stick to test the depth of the pool of water she was crossing (Breuer et al. 2005); the same individual used the trunk as a bridge to cross a deep section of swamp (Breuer et al. 2005). A female wild mountain gorilla was observed extending a bamboo pole to her infant who then climbed the pole so that it functioned as a ladder (Grueter et al. 2013). Orangutans were observed

who first bit through the bottoms of vines to free them and then used the vines to swing across a gap in the forest (van Schaik et al. 2006). A gorilla was observed throwing a detached branch toward researchers, and a group of gorillas threw fistfuls of grass at a man who threw stones at them (Wittiger & Sunderland-Groves 2007). Such behaviors, once observed for the first time, may subsequently become the focus of future ethograms if believed to possibly be recurrent, but are not included in the database compiled by Reader and colleagues (2011) and used in the present study, if described by the original observers as only a possible instance of tool use.

A few taxonomies of tool use have been proposed, though none is consensually used. Parker and Gibson (1977, 1979) proposed a taxonomy based on the motor aspects involved in the behavior, including simple prehension, simple object manipulation, object-substrate manipulation, complex object manipulation, and social-object manipulation. This classification derived from Piaget's (1952) theory of sensorimotor stages in human children. In contrast, Silva and Silva (2016) taxonomy distinguishes forms of tool use based on their function, and therefore the context of the behavior, rather than on motor aspects. It includes use of stick tools for foraging; the use of hammers and anvils for foraging; the use of a variety of tools (such as sponges made of moss or leaves, or husks) for fluid transportation; the use of tools for climbing or assisting other motion behaviors; and the use of weapons in agonistic encounters. This taxonomy takes into account the fact that species differ in how they achieve a goal, but their behaviors may serve similar purposes. For instance, in contrast to blonde capuchins who fish for termites above the ground, chimpanzees fish for this insect on the ground. Considering the lack of consensus on the taxonomic organization of tool use, and the fact that Reader and colleagues' (2011) interest lied primarily on the G factor (i.e., the highest level of the factor structure of complex problem solving) rather than on subtypes of each specific problem-solving ability, they

did not divide tool use behavior reports into subtypes, but rather on the catch-all “tool use” classification.

3. Social Learning

Social learning is defined as learning that is facilitated by observation of, or interaction with, another individual and/or the products of its behavior (Box, 1984; Galef, 1988). Social learners are not uncommonly characterized as *information scroungers*, as they exploit the information, skills, and resources obtained or devised through asocial learning (Laland, 2004). While intentional teaching may be included in human social learning, it is so rarely observed in non-human animals (and even so, dubious in most cases) that this animal literature definition does not include teaching. Rather, the social learning can usually happen:

a) through observation; starting with the acquisition of a trait by individual A, followed by performance of trait by individual A, with concomitant observation by individual B, and ultimately followed by performance of the trait by individual B; or less commonly
b) through exposure to products; starting with the acquisition of a trait by individual A, followed by the performance of trait by individual A, followed by the occurrence of product of the trait (e.g., an open nut), subsequent exposure of individual B to the product, and ultimately followed by performance of the trait by individual B (Hoppitt & Laland, 2013).

Note however, that the term “trait” is open-ended in the above definition, and thus such an operationalization is somewhat vague. In some cases, the trait is the capacity to interact with the *same kind of object*, irrespective of how; in others, the trait is the capacity to repeat the *behavior* irrespective of the target being the same object (Hoppitt & Laland, 2013). For social learning

with the highest fidelity however, the trait refers to the capacity to interact with the same kind of object, through the same behavior.

To organize the observation and understanding of social learning theoretically it is proposed that social learning involves the observations of (1) a specific kind of individual (the "who" aspect), (2) performing particular behaviors or acting upon particular targets (the "what" aspect), (3) under suitable circumstances (the "when" aspect) (Kendal et al., 2018). For instance, the "who" aspect might be preferentially kin, successful individuals, or older individuals. The "what" aspect might be behavior that is consistently emitted by kin but not yet by the social learner. The "when" aspect might be especially in situations in which the individual is uncertain, or when asocial learning is too costly, as in fact several primate species have demonstrated to socially learn primarily in more complex tasks (Kendal et al., 2018).

This theoretical proposal to guide the observation of social learning is too recent though, and until now relatively little effort has been put int investigating the form and basic mechanisms, rather than the functions, of social learning in non-human animals. Researchers have amassed large data banks containing the social diffusion of innovations, and even included geographical variation in many species, yet the contexts in which individuals tend to copy others remain largely unexplored (Heyes, 1994; Hoppitt & Laland, 2013; Laland, 2004). Already in 1994, Heyes identified that "it is rarely possible to assign an example of social learning to one category rather than another with certainty, and there is little consensus regarding the use of category labels", and as a consequence of the continued lack of study on how social learning occurs and is characterized, that statement is still true. As such, a taxonomy of kinds of social learning not possible, with all reported instances of it collated by Reader and colleagues (2011) being included in the same single category.

Upon ethological observation of behavior, the notion that "social learning" has happened is an assumption based on the temporal contiguity between noting that an individual observed a behavior and subsequently enacted the behavior, and the similarity between the behavior it observed and the one subsequently enacted. The following description of a methodology to record social learning instances in tufted capuchin monkeys, by Coelho and colleagues (2015, p. 913), illustrates the process:

> *[...] we adopted a protocol in which the focus was maintained on an ongoing event until it ended, simultaneous events being ignored. We named this method "Focal Event" as the focus was the activities surrounding a nut-cracking episode itself and not the monkey performing the behavior. A Focal Event began when a monkey approached an anvil and began to manipulate the elements there present and ended when it left the anvil or—in the event of a conspecific observer being present (after the capuchin manipulating the tool had left the anvil), when the observer also left the nut-cracking site. With this methodology, we could access not only observation and scrounging opportunities during a nut-cracking event but the subsequent actions of the conspecific observer, including delayed scrounging and manipulation of the elements within the nut-cracking site.*

4. Tactical Deception

The definition of tactical deception has been less variable in the literature, perhaps because of its being predominantly influenced, both in terms of numbers of publications and their impact, by a few authors: Richard Byrne, Andrew Whiten, and collaborators. Tactical deception involves "acts from the normal repertoire of an individual, used at low frequency and in contexts different from those in which it uses the high frequency (honest) version of the act,

such that another familiar individual is likely to misinterpret what the acts signify, to the advantage of the actor" (Byrne & Whiten 1985, p. 672). It is uncommon for deception to involve novel behavior, with a few exceptions having been observed in great apes. Rather, it involves behaviors already present in the repertoire of the individuals, but in a different context or with slight modifications (Byrne, 2003). "Tactical" therefore refers to this ability of shifting one aspect of one's behavioral repertoire into another (deceptive) role. One example involves a baboon suddenly taking an alert posture and shifting its body and gaze towards the horizon (a behavior normally exhibited when a predator or a threatening neighboring group is spotted), when being chased aggressively by a familiar conspecific (Whiten & Byrne, 1988). The pursuer stopped to look for a threat in the same direction, and never resumed the chase, though no real threat existed.

Instances of tactical deception are frequently observed being repeatedly enacted by the same individual (sometimes in quick succession, sometimes sparsed over many hours or days; Whiten & Byrne, 1988), sometimes on the same victim and other times the animal will have tried the tactic on other victims before; therefore instances of tactical deception are rarely behavioral innovations. Note that the main definition of tactical deception quoted above specifies that victims are "familiar individuals", to emphasize that the behaviors take place in a group of mutually recognizable individuals with frequent re-encounters; therefore, pressures exist to make the deceptive behaviors be complex, manipulative, and convincing as giving the appearance of non-malevolence, rather than simply being overt acts of desertion, exclusion, exploitation, appropriation, or thievery (Cheney & Seyfarth 1985; van Rhijn & Vodegel 1980).

Most tactical deception behaviors appear to be simply a product of associative learning (Byrne and Whiten, 1990; Byrne, 1997), whereby an individual primate associated the behavior

implementation with successful acquisition of rewards. However, records that seemed unlikely to be explained simply by associative learning were mostly of great apes (and not of monkeys or non-simian primates), with Byrne and Whiten (1992) therefore suggesting that they had some form of theory of mind, and the ability to anticipate deception and its potential outcomes.

Some categorizations of tactical deception behaviors into a taxonomy have been proposed, with the most detailed being that forwarded by Whiten & Byrne (1988). Categories proposed are emergent from data, and based on the inferred functional consequences of each deceptive act. Being data-driven, it is important to note that future ethological records may contribute to the identification of new classes of tactical deception; for instance, were examples of deception involving imitation of predators to be described, a new category to encompass that may emerge).

Five major functional forms of tactical deception were categorized by Whiten & Byrne (1988), each with subtypes:

1. Concealment (i.e., the agent conceals something from the target), including:

a) hiding (an object, part of self, or whole self) from view;

b) accoustic concealment; that is, acting quietly so as not to attract the target's attention;

c) inhibition of attending to a desirable object; that is, (temporarily) avoiding looking at or approaching something the agent desires so as not not to lead one or more targets to notice it.

As an example of concealment, among gelada baboons, individuals involved in extra-pair copulations are less likely to vocalize (le Roux et al., 2013).

2. Distraction; that is, the agent diverts the target's attention from some locus toward a second locus, so that the agent can subsequently focus on the first free of competition. This category includes:

a) distracting by looking away;

b) distracting by looking away with linked vocal signal;

c) distracting by leading away;

d) distracting by intimate behavior, including presenting, grooming, and exhibiting facial expressions.

As an example of distraction (Jolly, 1985): A male who did not usually share food items caught an antelope. A female who was known by the researchers to be particularly fond of meat approached him and groomed him for a while until he lolled back. As soon as that happened, she snatched the antelope carcass and ran away.

3. Creating an image of self; that is, affecting the way the target interprets the behavior(s) of the agent. This category includes:

a) presenting a neutral image; that is, present oneself as non-threatening;

b) presenting an affiliative image.

As an example of this class of deception (de Waal, 1982): A female chimpanzee was observed that, if unable to grab hold of her opponent successfully during a fight, would invite the opponent to reconciliation. She did that by holding out her hand, but when the other hesitantly put her hand in to agree to a reconciliation, the deceiver would suddenly and aggressively grab hold of her.

4. Manipulation of target individual using social tool (a third individual); that is, in this triadic relationship the agent's behavior is directed toward one individual but ultimately affects another through the mediating individual. This category includes:

a) deceiving tool about agent's involvement with target;

b) deceiving tool 1 about tool 2's involvement with target (here, four individuals are involved in total);

c) deceiving target about agent's involvement with social tool.

As an example (Byrne & Whiten 1985): An adult female chacma baboon was observed feeding in a specific patch, when a young juvenile tentatively approached and, though the adult female made no threat, the juvenile screamed. The only adult male of the group ran towards them and the adult female retreated, leaving the juvenile to feed on her food source.

5. Deflection of target to fall guy; that is, diverting an individual posing a problem, toward a passive victim. As an example (Whiten & Byrne, 1988): An adult male who was feeding was approached by a better-ranked male; when he noticed that he attacked a feeding female nearby briefly. The higher-ranked male then barked and chased her, leaving the former male feeding by himself.

Alternative categorizations of tactical deception have been made, such as by Byrne (2003). This categorization does not constitute a multi-level taxonomy, but rather a simpler grouping of behaviors into 11 single-level classes of behavior: exaggeration of behavior, ignoring signals, modifying one's own or another's body, suppression of signals, suppression of signs, hiding (part of) body or object, showing affiliation or neutral behavior, displaying or

calling at nothing or at neutral stimulus, leading others to/from something, use of "displaced" actions, and invention of actions. For a detailed description of these alternative, less commonly-used categories (relative to the Whiten and Byrne's 1988 taxonomy), see Byrne (2003). Importantly, note that most of the four other cognitive abilities have not been intricately decomposed to this degree. Because of this, and also because of the low numbers of clearly already-categorized instances of tactical deception into a specific taxonomy, and also because of their interest in higher levels of the hierarchical organization of *G* and not in specific abilities within, Reader and colleagues (2011) included only the broad level “tactical deception” in the database used in the present study. As the present study is equally interested in G and not in the intricacies of specialized abilities, the lack of reliance on taxonomic subdivisions of tactical deception imposes no limitations to the goals outlined above for the analyses that will follow here.

5. Innovation

Innovation is commonly defined as a behavioral solution that departs from the standard repertoire of behaviors of the species (Kummer & Goodall, 1985, 2003; Lefebvre et al., 1997). While it is difficult to assert what the “normal repertoire” is, especially in species with wide or flexible repertoires, some behaviors especially contrast with actions described before for the same species, or resemble only actions used in other contexts (Byrne, 2003). An innovation can thus not only be a novel solution to a novel problem, or a novel solution to an old one, but also an already-existing solution but applied to a different problem. Note that the term “invention” need not connote the notion that there was intention or planfulness in its development, as it may

arise through explorative behavior, experimentation, and associative learning (Ramsay et al., 2007).

Operationally, the range of behaviors that can be classified as innovative is wide, as an innovation can arise in any area of problem-solving (feeding, communicating, mating, parenting, agonistic interactions, etc.). It is difficult to identify innovations in the wild partly because many are progressively incorporated into the behavioral repertoire of the community – potato washing in Japanese macaques, for example, spread through social learning rather quickly (Kawai, 1965). Therefore, if the initial, seminal invention had not been observed and registered, the subsequent somewhat widespread implementation of potato washing would not be perceived as the product of innovation, but rather as a common, at most cultural aspect. Innovations in fact commonly spread socially, including not only potato washing, but also wheat sluicing by Japanese macaques (in fact both invented by the same individual; van Schaik, 2003).

Experience and familiarity with the components of the situation seem to be associated with higher innovation rates (Kummer & Goodall, 2003). Individual differences in personality, such as in boldness and proactivity, appears to affect responses to novel objects, affecting innovation. Species differences in the tendency to explore novel objects also exist (Glickman & Sroges, 1966), perhaps affecting innovation rates.

It is possible that focusing on exploration and curiosity, rather than on innovations per se, would ease the difficulty of operationalizing the observations and thus facilitate the identification of relevant behaviors. However, Reader and colleagues (2011) chose to keep with the decision of collation innovation instances.

Some more detailed examples of innovation are listed below:

a) A male chimpanzee was observed keeping somewhat close to an attractive female during the night, rising before other males then signaling to her in order to lead her away before others would notice (Tutin, 1979).
b) An older female chimpanzee was observed terminating the provocative, challenging and aggressive behavior of an adolescent male by tickling him. Usually such behaviors provoke either fear and avoidance, or retaliation (Kummer & Goodall, 1985)
c) A female chimpanzee was observed sitting in between two aggressive, disputing males to facilitate their reconciliation: both would groom her and she would subsequently step quietly away and leave them grooming each other (Kummer & Goodall, 1985).
d) An infant chimpanzee was observed using the same approach commonly used to open hardshelled Strychnos fruits with a rock, to smash an insect possessing a hard exoskeleton (Kummer & Goodall, 1985).
e) A tufted capuchin monkeys were observed, for the first time in its community, washing sandy food, and when the behavior spread by social learning, another individual developed the unusual strategy of stealing only already-washed food rather than learning how to clean it by itself (Visalberghi & Fragasy, 1990).

As can be readily noticed, there is no consistent pattern across instances of innovation recorded in the literature, making it difficult to implement a specific and clear ethogram to identify novel innovations. Rather, it is by noting unusual problem-solving behavior that such innovations are identified

Appendix A References

Aguiar, L. M., Cardoso, R. M., Back, J. P., Carneiro, E. C., Suzin, A., & Ottoni, E. B. (2014). Tool use in urban populations of capuchin monkeys Sapajus spp.(Primates: Cebidae). *Zoologia (Curitiba)*, *31*(5), 516-519.

Amant, R., & Horton, T. E. (2008). Revisiting the definition of animal tool use. *Animal Behaviour*, *75*, 1199-1208.

Beck, B. B. (1980). The use and manufacture of tools by animals. *Garland STPM Press.*

Box, H. O. (1984). *Primate behavior and social ecology*. Chapman & Hall.

Breuer, T., Ndoundou-Hockemba, M., & Fishlock, V. (2005). First observation of tool use in wild gorillas. *PLoS Biology*, *3*, e380.

Byrne, R. W. (1995). Primate cognition: comparing problems and skills. *American Journal of Primatology*, 37, 127–141.

Byrne, R. W. (1997). Machiavellian intelligence. Evolutionary Anthropology, 5, 172–180

Byrne, R. W. (2003). Novelty in deceit. In S. M. Reader & K. N. Laland (Eds.), *Animal innovation* (pp 237-260). Oxford University Press.

Byrne, R. W., & Whiten, A. (1985). Tactical deception of familiar individuals in baboons (Papio ursinus). *Animal Behaviour*, *33*, 669-673.

Byrne, R. W., & Whiten, A. (1992). Cognitive evolution in primates: evidence from tactical deception. *Man*, 609-627.

Byrne, R. W., Whiten, A., & Henzi, S. P. (1990). Social relationships of mountain baboons: Leadership and affiliation in a non-female-bonded monkey. *American Journal of Primatology*, *20*, 313-329.

Cheney, D. L., & Seyfarth, R. M. (1985). Vervet monkey alarm calls: Manipulation through shared information?. *Behaviour*, *94*, 150-166.

Coelho, C. G., Falotico, T., Izar, P., Mannu, M., Resende, B. D. D., Siqueira, J. O., & Ottoni, E. B. (2015). Social learning strategies for nut-cracking by tufted capuchin monkeys (Sapajus spp.). *Animal Cognition*, *18*, 911-919.

Day, R. L., Coe, R. L., Kendal, J. R., & Laland, K. N. (2003). Neophilia, innovation and social learning: a study of intergeneric differences in callitrichid monkeys. *Animal Behaviour, 65*, 559-571.

Fragaszy, D., Izar, P., Visalberghi, E., Ottoni, E. B., & de Oliveira, M. G. (2004). Wild capuchin monkeys (Cebus libidinosus) use anvils and stone pounding tools. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *64*, 359-366.

Fragaszy, D., & Perry, S. (2003). *The biology of traditions: Models and evidence*. Cambridge University Press.

Galef, B. G. (1976). Social transmission of acquired behavior: a discussion of tradition and social learning in vertebrates. In: J. S. Rosenblatt, R. A. Hinde, E. Shaw & C. Beer (Eds.), *Advances in the Study of Behavior, vol. 6* (pp. 77-100). Academic Press.

Gibson, K. R., 1986. Cognition, brain size and the extraction of embedded food resources. In: Lee, E. (Ed.), *Primate ontogeny, cognition and social behaviour* (pp. 93-103). Cambridge University Press.

Glickman, S. E., & Sroges, R. W. (1966). Curiosity in zoo animals. *Behaviour*, *26*, 151-187.

Grueter, C. C., Robbins, M. M., Ndagijimana, F., & Stoinski, T. S. (2013). Possible tool use in a mountain gorilla. *Behavioural Processes, 100*, 160-162.

Gumert, M. D., Kluck, M., & Malaivijitnond, S. (2009). The physical characteristics and usage patterns of stone axe and pounding hammers used by long-tailed macaques in the Andaman Sea region of Thailand. *American Journal of Primatology*, *71*, 594-608.

Heyes, C. M. (1994). Social learning in animals: categories and mechanisms. *Biological Reviews*, *69*, 207-231.

Hoppitt, W., & Laland, K. N. (2013). *Social learning: an introduction to mechanisms, methods, and models*. Princeton University Press.

Jolly, A. (1985). The Evolution of Primate Behavior: A survey of the primate order traces the progressive development of intelligence as a way of life. *American Scientist*, *73*, 230-239.

Kawai, M. 1965. On the system of social ranks in a natural troop of Japanese monkeys. 1. Basic rank and dependent rank. In K. Imanishi & S. A. Altmann (Eds.) *Japanese Monkeys* (pp. 66–86). Emory University Press.

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences*, *22*, 651-665.

Kummer, H., & Goodall, J. (1985). Conditions of innovative behaviour in primates. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*, 203-214.

Kummer, H., & Goodall, J. (2003). Conditions of innovative behaviour in primates. In S. M. Reader & K. N. Laland (Eds.), *Animal innovation* (pp. 223–235). New York, NY: Oxford University Press

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*, 4-14.

van Lawick-Goodall, J. (1968). The behaviour of free-living chimpanzees in the Gombe Stream Reserve. *Animal Behaviour Monographs*, *1*, 161-IN12.

Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*, *53*, 549-560.

Le Roux, A., Snyder-Mackler, N., Roberts, E. K., Beehner, J. C., & Bergman, T. J. (2013). Evidence for tactical concealment in a wild primate. *Nature Communications*, *4*, 1462.

Melin, A. D., Young, H. C., Mosdossy, K. N., & Fedigan, L. M. (2014). Seasonality, extractive foraging and the evolution of primate sensorimotor intelligence. *Journal of Human Evolution*, *71*, 77-86.

Parker, S. T., & Gibson, K. R. (1977). Object manipulation, tool use and sensorimotor intelligence as feeding adaptations in Cebus monkeys and great apes. *Journal of Human Evolution*, *6*, 623-641.

Parker, S. T., & Gibson, K. R. (1979). A developmental model for the evolution of language and intelligence in early hominids. *The Behavioral and Brain Sciences*, *2*, 367–408.

Piaget, J. (1952). *The child's conception of number*. Routledge & Kegan Paul.

Ramsey, G., Bastian, M. L., & van Schaik, C. (2007). Animal innovation defined and operationalized. *Behavioral and Brain Sciences*, *30*, 393-407.

Reader, S. (2003). Innovation and social learning: individual variation and brain evolution. *Animal Biology*, *53*, 147-158.

Reader, S. M., & Laland, K. N. (2001). Primate innovation: sex, age and social rank differences. *International Journal of Primatology*, *22*, 787-805.

Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1017-1027.

Van Rhijn, J. G., & Vodegel, R. (1980). Being honest about one's intentions: an evolutionary stable strategy for animal conflicts. *Journal of Theoretical Biology*, *85*, 623-641.

van Schaik, C. P., Fox, E. A., & Sitompul, A. F. (1996). Manufacture and use of tools in wild Sumatran orangutans. *Naturwissenschaften*, *83*, 186-188.

van Schaik, C. P., Fragaszy, D., & Perry, S. (2003). Local traditions in orangutans and chimpanzees: social learning and social tolerance. *The biology of traditions: Models and evidence*, 297-328.

van Schaik, C. P., van Noordwijk, M. A., & Wich, S. A. (2006). Innovation in wild Bornean orangutans (*Pongo pygmaeus wurmbii*). *Behaviour, 143*, 839–876.

Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal tool behavior: The use and manufacture of tools by animals*. JHU Press.

Silva F.J., & Silva K.M. (2016). Primate tool use. In: Weekes-Shackelford V., Shackelford T., Weekes-Shackelford V. (eds) *Encyclopedia of Evolutionary Psychological Science*. Springer.

Souto, A., Bione, C. B., Bastos, M., Bezerra, B. M., Fragaszy, D., & Schiel, N. (2011). Critically endangered blonde capuchins fish for termites and use new techniques to accomplish the task. *Biology Letters*, *7*, 532-535.

Tutin, C. E. (1979). Mating patterns and reproductive strategies in a community of wild chimpanzees (*Pan troglodytes schweinfurthii*). *Behavioral Ecology and Sociobiology*, *6*, 29-38.

Visalberghi, E., and Fragaszy, D. (1990). Do monkeys ape? In Parker, S., and Gibson, K. (eds.), *Comparative developmental psychology of language and intelligence in primates* (pp. 247-273). Cambridge University Press.

de Waal, F. (1982). *Chimpanzee politics*. Jonathan Cape.

Whiten, A., & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences*, *11*, 233-244.

Wittiger, L., & Sunderland-Groves, J. L. (2007). Tool use during display behavior in wild Cross River gorillas. *American Journal of Primatology: Official Journal of the American Society of Primatologists, 69*, 1307-1311.

Xiang, Z., Liang, W., Nie, S., & Li, M. (2013). A short note on extractive foraging behavior in gray snub-nosed monkeys. *Integrative Zoology*, *8*, 389-394.

Appendix B: Supplementary Analyses to Chapter 1

1. Confirmatory replication with a set of overlapping datasets

Many previous examinations of the evolutionary patterns behind primate traits using cross-species data relied on datasets of diverse sizes and compositions (e.g., Blomberg et al., 2003; Cooper & Purvis, 2010; Deaner et al., 2006, Kamilar & Cooper, 2013), each drawing conclusions by comparing only partly overlapping databases. This is somewhat justified by fact that species within taxonomic groups in primates have largely shared their history and partly share ecological conditions and thus adaptations as well, thus serving, to some extent, interchangeably in "samples" of primate species from which ancestral characteristics can be estimated. Nevertheless, species are clearly not perfectly representative of each other, especially in traits that exhibit evolutionary lability rather than conservativeness. To examine whether the results identified in Chapter 1 hold when including only species present in all databases, all phylogenetic coefficients were re-estimated for the set of species for which all datasets overlap. The resulting database contained 32 species. All the same analytical steps were followed as in the original paper, resulting in the estimates displayed on Table A2.1.

Table A2.1.

Parameter estimates for phylogenetic conservatism, evolutionary rate, rate acceleration, early burst, and selection regime strength, for the perfectly overlapping sample of 32 primate species.

Measure	$\boldsymbol{\lambda}$	$\boldsymbol{K}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\delta}$	$\boldsymbol{a}$	$\boldsymbol{\alpha}$
G	.722	.264	.063	10.500	.000	.075
Residual brain	.999	.598	.035	1.816	.000	.006

Residual neocortex	.902	.487	.043	3.228	.000	.023
Residual cerebellum	.699	.331	.050	7.631	.000	.059
Neocortex ratio	.999	1.360	.012	.749	.000	.000
Brain size	.999	1.589	.013	.740	-.024	.000
Neocortex size	.999	1.352	.016	.946	-.010	.000
Hippocampus size	.999	.761	.022	1.600	.000	.004
Cerebellum size	.999	1.344	.014	.779	-.029	.000
Body size	.999	1.425	.013	.917	-.016	.000
Correlation (*r*) with full dataset estimates	.850	.928	.846	.962	.948	.827

It can be observed that the comparability with the original results is high as:

1) *G* and residual cerebellum size exhibit the lowest phylogenetic conservatism as measured by λ and *K*, followed by residual neocortex size. Most other variables conformed to the Brownian motion expectation when examining λ, or were even more conserved than expected in Brownian motion, just as in the original analyses;
2) *G* exhibited the fastest evolutionary rates, again followed by the residual cerebellum and residual neocortex size measures.
3) *G* exhibited the strongest evolutionary rate acceleration as measured by $\boldsymbol{\delta}$, and strongest selection regime towards an optimum, as measured by α, similarly followed by residual cerebellum size, and to a lesser degree residual neocortex size.
4) Non-residualized NVMs exhibited negative a values, indicating an early burst of changes, but subsequent deceleration of rates, just as in the original results.

In line with these observations, correlations between the original estimates and the estimates (examined for each of the columns in Table A2.1, and displayed at the last row) ranged from r = 827 to .962 ($p < .05$ for all indices). While these are not perfect correlations, they indicate largely converging patterns. The main differences reside in the *magnitude* of some estimates, although the rankings remain the same as in the original results. Notably, evolutionary rates, acceleration estimates, and selection regime strength estimates were lower than in the original analyses, consistently across measures. This may indicate that the high rates, acceleration, and strength of regimes observed in the original analyses were partly driven by some taxa removed in the present replication. As mentioned in Chapter 1 however, clade-specific analyses are one of the future steps already in preliminary implementation, to examine how different primate taxa contrast in the phylogenetic patterns described in that Chapter and in this Appendix section.

2. Fit of evolutionary models

Examining the relative fit of models can permit identifying to which the data *best* conform, and which models are poor representatives of the real data. This can be informative in the case of Chapter 1, where the following models are tested for each measure: acceleration, early burst, Ornstein-Uhlenbeck, and phylogenetic signal. However, considering that several of these models can be complementary (e.g., acceleration may go hand in hand with the Ornstein-Uhlenbeck expectation of a strong selection regime toward an optimal value), it is not necessarily the case that one needs to be selected while others are discarded. As such, as these models are not alternative, contrasting their fit statistically is not strictly needed. Still, displaying their relative fit is a transparent demonstration of how well the models explain the data.

The Akaike Information Criterion is a statistical fit index that takes into consideration the log likelihood of a model for a particular variable, and penalizes more complex models (i.e., those with more parameters being estimated) if they do not add explanatory power (Hurvich, Simonoff, & Tsai, 1998). Thus AIC is ideal to select parsimonious models. No absolute cutoffs exist (e.g., 0 or 1), but AIC can be used to compare the fit of different models on the same dataset, as lower AIC values reflect better fit. For small datasets, AICc is used to correct for possible biases.

Weights associated to each Akaike Information Criteria (AICc) value can be used to determine the best model (Wagenmakers & Farrell, 2004). Instead of subjectively comparing AIC values across models, they can be transformed into *Akaike weights* (for a detailed review and description of this approach, see Wagenmakers & Farrell, 2004), which can be interpreted as conditional probabilities for each model.

AICc and weights (AICc w_i) are displayed in Table A2.2. It can be observed that G and residual cerebellum size were best explained by the same models (acceleration and the Ornstein-Uhlenbeck model of strong selection regime towards an optimum). Other residual measures were better explained by phylogenetic signal (i.e., evolutionary lability), suggesting that while they are not conserved as would be expected under Brownian motion, they are lability in not so consistent directions and not so much in an accelerating way. Finally, absolute measures conformed either to early-burst (suggesting they decelerate over time in terms of their evolutionary rate of change) or to Brownian motion, indicating that they are phylogenetically conserved variables. These results compound the results outlined in Chapter 1, complementing rather than contrasting.

Table A2.2.

Absolute (AICc) and relative (AICc w_i) statistical fit of evolutionary models, along with the favored model for each trait based on AICc w_i

Measure	Brownian motion	Acceleration	Early burst	Ornstein-Uhlenbeck	Phylogenetic signal	Favored model
	AICc (AICc w_i)	AICc (AICc w_i)	AICc (AICc w_i)	AICc (AICc w_i)	AICc (AICc w_i)	
G factor	176.94 (<.01)	131.51 (.36)	179.13 (.00)	131.49 (.36)	132.38 (.27)	δ / OU
Residual brain	333.01 (<.01)	321.25 (<.01)	335.08 (<.01)	319.01 (<.01)	287.031 (.99)	PS
Residual neocortex	61.05 (<.01)	51.95 (.21)	62.91 (<.01)	51.20 (.23)	49.40 (.56)	PS
Residual cerebellum	62.04 (<.01)	28.93 (.41)	64.23 (<.01)	28.75 (.44)	30.88 (.15)	δ / OU
Neocortex ratio	27.84 (.29)	29.43 (.13)	29.95 (.10)	29.41 (.13)	27.54 (.34)	PS / BM
Absolute brain size	83.63 (<.01)	76.92 (<.01)	60.38 (.99)	85.71 (<.01)	85.66 (<.01)	EB
Absolute neocortex size	120.28 (.42)	122.41 (.15)	122.45 (.14)	122.38 (.15)	122.52 (.14)	BM
Absolute hippocampus size	99.04 (.36)	100.19 (.20)	100.96 (.14)	100.76 (.15)	100.84 (.15)	BM
Absolute cerebellum size	123.41 (.21)	124.87 (.10)	121.55 (.54)	125.61 (.07)	125.61 (.07)	EB
Absolute body mass	226.91 (.07)	226.61 (.09)	222.42 (.69)	227.89 (.05)	226.24 (.10)	EB

3. Interpretability of evolutionary rates

The interpretability of comparing several measures in terms of evolutionary rates can be influenced by how much absolute variation they exhibit (i.e, the size of their standard deviations relative to their respective mean). Take *i* and *j* as two hypothetical neuroanatomical volume measures. If the vector of scores on *i* is [1001, 1002, 1003, 1004, 1005] and if the vector of scores for *j* is [1000, 2000, 3000, 4000, 5000], *i* has a much lower coefficient of variation than *j*. It is easier to evolve from 1001 to 1002 than from 1000 to 2000. While *z*-scores are the same in both hypothetical vectors, *j* exhibits more variation relative to the mean. In other words, the scores are more dispersed among themselves relative to the mean, in spite of *z*-scores for both vectors being equivalent. If *i* were to exhibit a higher standardized evolutionary rate than *j*, that would unfortunately not mean that *i* evolves faster in *absolute* terms, because a faster evolutionary rate for *i* compared to *j* would simply mean that species' scores change from 1001 to 1002, to 1003, etc. faster on *i* than they do from 1000 to 2000, to 3000, etc. on *j*.

The coefficient of variation on variables compared in terms of their standardized evolutionary rates (i.e., evolutionary rates based on z-scores) is especially informative in the context of coefficients of variation. The coefficient of variation of a variable equates to its standard deviation divided by its mean, reflecting how much relative variation exists in it. Variables that exhibit both (a) higher evolutionary rates than others and (b) either comparable or higher coefficient of variation can be interpreted as truly evolving faster in absolute terms than others. As such, care was taken to examine if variables presented in Figure 1 which exhibited higher evolutionary rates had comparable or higher coefficients of variation compared to others. The coefficient of variation for all neuroanatomical measures ranged from .14 to .16, being thus highly comparable among themselves. *G*, which exhibited higher evolutionary rates than

neuroanatomical volume measures, exhibited a coefficient of variation of .36. It can be inferred that its faster evolutionary rates compared to volume measures involve even more absolute change across evolutionary time than if *G* and volume measures had equivalent coefficients of variation. As such, it is possible that the estimate of evolutionary rate of *G* presented in Chapter 1 is somewhat underestimated, though any assertive statement or correction is unlikely to be precise considering that the relationship of coefficient of comparability to standardized evolutionary rates is still not well known (Adams, 2013).

Appendix B References

Adams, D. C. (2013). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology, 62*, 181–192.

Blomberg, S. P., Garland Jr, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, *57*, 717-745.

Cooper, N., & Purvis, A. (2010). Body size evolution in mammals: complexity in tempo and mode. *The American Naturalist*, *175*, 727-738.

Deaner, R. O., Van Schaik, C. P., & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology*, *4*, 149-196.

Hurvich, C. M., Simonoff, J. S., & Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *60*, 271-293.

Kamilar, J. M., & Cooper, N. (2013). Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*, 20120341.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192-196.

Appendix C: Supplementary Analyses to Chapter 3

1. Reanalyses with only five highly correlated telencephalic regions

Although located in the telecephalon, the volumes of the striatum, the bulbus olfactorius and the bulbus olfactorius accessorius exhibited low (and in some designs, non-significant) factor loadings in the common factor among telencephalic regions, in Chapter 3. Considering the possibility that their inclusion in all analyses involving regions of the telencephalon in spite of their lack of important association with other regions in terms of their sizes across species, analyses of anatomical integration for the subregions of the telencephalon were fully repeated without their inclusion. Thus the only included regions were the neocortex, hippocampus, lobus piriformis, septum, and schizocortex.

Table A3.1 displays results of analyses without body size controls. Factor loadings were similar in strength to those observed when the striatum, the bulbus olfactorius and the bulbus olfactorius accessorius were included, both without and with phylogenetic controls. Although all were positive, anatomical integration coefficients were small and nonsignificant.

Table A3.1.

Average factor loadings and anatomical integration indices for regions of the primate telencephalon, without and with phylogenetic control

Brain region	Factor loading	Evolutionary factor loading	Anatomical integration coefficient	Evolutionary anatomical integration coefficient
Neocortex	.999*	.994*	.210	.151

Hippocampus	.928*	.911*	.147	.228
Lobus piriformis	.965*	.966*	.210	.216
Septum	.981*	.975*	.216	.230
Schizocortex	.949*	.948*	.196	.258

Note: * $p < .05$.

Table A3.2., in turn, displays results after body size is controlled for, in all variables. Factor loadings were overall somewhat slightly stronger than those observed for the equivalent analysis in Chapter 3. It can be observed that telencephalon size positively predicted the strength of the relationship of hippocampal size to the size of other regions.

Table A3.2.

Average factor loadings and integration indices for regions of the primate telencephalon residualized against body size, without and with phylogenetic control

Brain region	**Factor loading**	**Evolutionary factor loading**	**Anatomical integration coefficient**	**Evolutionary anatomical integration coefficient**
Neocortex	.984*	.985*	.223	.188
Hippocampus	.961*	.685*	.326*	.197
Lobus piriformis	.970*	.861*	.265	.296
Septum	.989*	.883*	.244	.292
Schizocortex	.980*	.759*	.242	.219

Note: * $p < .05$.

Closer inspection revealed that residual telencephalic size positively and significantly ($p < .05$) predicted the relationship of residual hippocampal size to the size of each and every other telencephalic region ($r = .297$, $r = .368$, $r = .322$, $r = .298$ for neocortex, lobus piriformis, septum, and schizocortex size). However, no integration of residual hippocampal size with the residual size of other regions as s function of residual telencephalon size was observed once phylogenetic control was conducted.

These complementary analyses suggest that, barring a possible integration of hippocampal size with the size of other regions as a function of telencephalon size, the original results were strongly replicated, both in factor loadings and in the overall lack of anatomical integration, lending validity to the analyses presented in Chapter 3. This indicates that the size of the telencephalon does not generally predict whether the evolution of its components undergoes more strongly mosaic or more strongly concerted evolution; rather, the degree of correlation among telencephalon components is largely the same across taxa of different telencephalic sizes.

References

Adams, D. C. (2013). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology, 62*, 181–192.

Arbilly, M., & Laland, K. N. (2017). The magnitude of innovation and its evolution in social animals. *Proceedings of the Royal Society B: Biological Sciences, 284*, 20162385.

Arden, R., & Zietsch, B. P. (2017). An all-positive correlation matrix is not evidence of domain-general intelligence. *Behavioral and Brain Sciences, 40*, e197-e197.

Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: a new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews, 19*, 114-118.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological Review, 113*, 628.

Barrickman, N. L., Bastian, M. L., Isler, K., & van Schaik, C. P. (2008). Life history costs and benefits of encephalization: a comparative test using data from long-term studies of primates in the wild. *Journal of Human Evolution, 54*, 568-590.

Barton, R. A. (1996). Neocortex Size and Behavioural Ecology in Primates. *Proceedings of the Royal Society of London Series B, 263*, 173-177.

Barton, R. A. (2001). The coordinated structure of mosaic brain evolution. *Behavioral and Brain Sciences, 24*, 281-282.

Barton, R. A. (2006). Primate brain evolution: integrating comparative, neurophysiological, and ethological data. *Evolutionary Anthropology: Issues, News, and Reviews, 15*, 224-236.

Barton, R. A. (2012). Embodied cognitive evolution and the cerebellum. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*, 2097-2107.

Barton, R. A., & Harvey, P. H. (2000). Mosaic evolution of brain structure in mammals. *Nature, 405*, 1055.

Barton, R. A., & Venditti, C. (2013). Human frontal lobes are not relatively large. *Proceedings of the National Academy of Sciences, 110*, 9001-9006.

Barton, R. A., & Venditti, C. (2014). Rapid evolution of the cerebellum in humans and other great apes. *Current Biology, 24*(20), 2440-2444.

Blair, C. (2006) How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences 29*:109–25; discussion 125-60.

Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution, 57*, 717-745.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods, 10*, 689-709.

Boire, D., Nicolakakis, N., & Lefebvre, L. (2002). Tools and brains in birds. *Behaviour, 139*, 939-973.

Briffa, M., & Weiss, A. (2010). Animal personality. *Current Biology, 20*, R912-R914.

Burkart, J. M., Schubiger, M. N., van Schaik, C. P. (2017). Future directions for studying the evolution of general intelligence. *Behavioral and Brain Sciences, 40*, e224-e224.

Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 5*, 172-180.

Byrne, R. W. (2003). Novelty in deceit. In S. M. Reader & K. N. Laland (Eds.), *Animal innovation* (pp 237-260). Oxford University Press.

Byrne, R. W., & Corp, N. (2004). Neocortex size predicts deception rate in primates. *Proceedings of the Royal Society B: Biological Sciences*, *271*, 1693.

Byrne, R. W., & Whiten, A. (1985). Tactical deception of familiar individuals in baboons (*Papio ursinus*). *Animal Behaviour*, *33*, 669-673.

Byrne, R. W., & Whiten, A. (1988). Toward the next generation in data quality: a new survey of primate tactical deception. *Behavioral and Brain Sciences*, *11*, 267-273.

Cammarota, M., Bevilaqua, L. R. M., Köhler, C., Medina, J. H., &Izquierdo, I. (2005). Learning twice is different from learning once and from learning more. *Neuroscience*, *132*, 273-279.

Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist*, *29*, 512.

Chittka, L., & Niven, J. (2009). Are bigger brains better?. *Current biology*, *19*(21), R995-R1008.

Clark, D. A., Mitra, P. P., & Wang, S. S. (2001). Scalable architecture in mammalian brains. *Nature*, *411*(6834), 189-193.

Coelho Jr, A. M., & Bramblett, C. A. (1981). Interobserver agreement on a molecular ethogram of the genus Papio. *Animal Behaviour*, *29*, 443-448.

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the *g*-factor of intelligence. *Neuroimage*, *31*, 1359-1365.

Cooper, N., & Purvis, A. (2010). Body size evolution in mammals: complexity in tempo and mode. *The American Naturalist*, *175*, 727-738.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray.

Day, R. L., Coe, R. L., Kendal, J. R., & Laland, K. N. (2003). Neophilia, innovation and social learning: A study of intergeneric differences in callitrichid monkeys. *Animal Behaviour*, *65*, 559–571.

de Waal, F. (2016). *Are we smart enough to know how smart animals are?*. New York: Norton.

Deaner, R. O., Isler, K., Burkart, J., & van Schaik, C. (2007). Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, Behavior and Evolution*, *70*, 115-124.

Deaner, R. O., Nunn, C. L., & van Schaik, C. P. (2000). Comparative tests of primate cognition: different scaling methods produce different results. *Brain, Behavior and Evolution*, *55*, 44-52.

Deaner, R. O., van Schaik, C. P., & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology*, *4*, 147470490600400114.

Dicke, U., & Roth, G. (2016). Neuronal factors determining high intelligence. *Phil. Trans. R. Soc. B*, *371*, 20150180.

Doyon, J., Penhune, V., & Ungerleider, L. G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, *41*, 252-262.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of human evolution*, *22*, 469-493.

Dunbar, R. I. M, & Shultz, S. (2007). Evolution in the social brain. *science*, *317*(5843), 1344-1347.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., ... & Emslie, H. (2000). A neural basis for general intelligence. *Science*, *289*, 457-460.

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science*, *306*, 1903-1907.

Enard, W. (2014). Comparative genomics of brain size evolution. *Frontiers in Human Neuroscience*, *8*, 345.

Falk, D., & Gibson, K. R. (Eds.). (2001). *Evolutionary anatomy of the primate cerebral cortex*. Cambridge, UK: Cambridge University Press.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*, 783-791.

Felsenstein, J. (1988). Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, *19*, 445-471.

Fernandes, H. B. F. (2014). *Phylogenetic signal in carnivoran anatomy, life history, neurology, and ecology: How strong is it, and does allometry make a difference?*. Poster, Evolution 2014, Raleigh, NC.

Fernandes, H. B. F., Woodley, M. A., & te Nijenhuis, J. (2014). Differences in cognitive abilities among primates are concentrated on *G*: Phenotypic and phylogenetic comparisons with two meta-analytical databases. *Intelligence*, *46*, 311-322.

Fernandes, H. B. F., Peñaherrera-Aguirre, M., Woodley of Menie, M. A., & Figueredo, A. J. (In preparation). *The evolution of intelligence in primates: Differential trends among clades*.

Figueredo, A. J., Cabeza de Baca, T., Fernandes, H. B. F., Black, C. J., Peñaherrera, M., Hertler, S., ... & Meisenberg, G. (2017). A sequential canonical cascade model of social biogeography: Plants, parasites, and people. *Evolutionary Psychological Science, 3*(1), 40-61.

Figueredo, A. J., Cox, R. L., & Rhine, R. J. (1995). A generalizability analysis of subjective personality assessments in the stumptail macaque and the zebra finch. *Multivariate Behavioral Research, 30*, 167-197.

Figuerero, A. J., Ross, D. M., & Petrinovich, L. (1992). The Quantitative Ethology of the Zebra Finch: A Study in Comparitive Psychometrics. *Multivariate Behavioral Research, 27*, 435-458.

Figueredo, A. J., Woodley, M. A., Brown, S. D., & Ross, K. C. (2013). Multiple successful tests of the Strategic Differentiation-Integration Effort (SD-IE) hypothesis. *Journal of Social, Evolutionary, and Cultural Psychology*, *7*, 361-383.

Fingelkurts, A. A., Fingelkurts, A. A., & Kähkönen, S. (2005). Functional connectivity in the brain—is it an elusive concept?. *Neuroscience & Biobehavioral Reviews, 28*, 827-836.

Finlay, B. L., Darlington, R. B., & Nicastro, N. (2001). Developmental structure in brain evolution. *Behavioral and Brain Sciences*, *24*, 263-278.

Fitzpatrick, J. L., Almbro, M., Gonzalez-Voyer, A., Hamada, S., Pennington, C., Scanlan, J., &Kolm, N. (2012). Sexual selection uncouples the evolution of brain and body size in pinnipeds. *Journal of Evolutionary Biology, 25*, 1321-1330.

Fleagle, J. G. (1985). Size and adaptation in primates. In Jungers W. L. (Ed.), *Size and scaling in primate biology* (pp. 1-19). New York: Plenum Press.

Forss, S. I., Willems, E., Call, J., & van Schaik, C. P. (2016). Cognitive differences between orang-utan species: a test of the cultural intelligence hypothesis. *Scientific Reports, 6*, 30516.

Frahm, H. D., Stephan, H., & Stephan, M. (1982). Comparison of brain structure volumes in Insectivora and Primates. I. Neocortex. *Journal fur Hirnforschung*, *23*, 375-389.

Freckleton, R. P. (2009). The seven deadly sins of comparative analysis. *Journal of Evolutionary Biology*, *22*, 1367-1375.

Galsworthy, M. J., Arden, R., & Chabris, C. F. (2014). Animal models of general cognitive ability for genetic research into cognitive functioning. In *Behavior genetics of cognition across the lifespan* (pp. 257-278). Springer, New York, NY.

Gangestad, S. W. (2011). Evolutionary processes explaining the genetic variance in personality: An exploration of scenarios. In D. M. Buss, & P. H. Hawley (Eds.), *The evolution of personality and individual differences* (pp. 338-375). Oxford: Oxford University Press

Garamszegi, L. Z. (Ed.). (2014). *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice*. New York: Springer.

Gibson, K. R. (1986). Cognition, brain size and the extraction of embedded food resources. In J. G. Else (Ed.), *Primate evolution* (pp. 95–103). New York: Cambridge University Press.

Gibson, K. R. (1991). Myelination and behavioral development: A comparative perspective on questions of neoteny, altriciality, and intelligence. In K. R. Gibson & A. C. Petersen (Eds.), *Brain maturation and cognitive development* (pp. 29–63). New York: Aldine de Gruyter.

Gibson, K. R., & Ingold, T. (Eds.). (1993). Tools, language, and cognition in human evolution. Cambridge: Cambridge University Press.

Gibson, K. R., Rumbaugh, D. M.,& Beran, M. J. (2001). Bigger is better: primate brain size in relationship to cognition. In D. Falk and K. R. Gibson (Eds.), *Evolutionary anatomy of the primate cerebral cortex* (pp. 79–97). Cambridge, UK: Cambridge University Press.

Gingerich, P. D. (2009). Rates of evolution. *Annual Review of Ecology, Evolution, and Systematics, 40*, 657–675.

Girotto, V., & Tentori, K. (2008). Is domain-general thinking a domain-specific adaptation?. *Mind & Society*, *7*, 167-175.

Glickman, S. E., & Sroges, R. W. (1966). Curiosity in zoo animals. *Behaviour*, *26*, 151-187.

Gómez-Robles, A., Hopkins, W. D., & Sherwood, C. C. (2014). Modular structure facilitates mosaic evolution of the brain in chimpanzees and humans. *Nature communications*, *5*, 4469.

Gonzalez-Voyer, A., Winberg, S., & Kolm, N. (2009). Distinct evolutionary patterns of brain and body size during adaptive radiation. *Evolution*, *63*, 2266-2274.

Gorsuch, R. L. (1983). *Factor analysis*. Hillside, NJ: L. Erlbaum.

Gorsuch, R. L. (2005). Continuous parameter estimation model: Expanding the standard 26 statistical paradigm. *Journal of the Science Faculty of Chiang Mai University*, *32*, 11-27.

Gould, S. J. (1977). *Ontogeny and phylogeny*. Belknap Press of Harvard University Press.

Hager, R., Lu, L., Rosen, G. D., & Williams, R. W. (2012). Genetic architecture supports mosaic brain evolution and independent brain–body size regulation. *Nature communications*, *3*, 1079.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2004). Structural brain variation and general intelligence. *Neuroimage*, *23*, 425-433.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: sex matters. *NeuroImage*, *25*(1), 320-327.

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*, *51*, 1341-1351.

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., ... & Purvis, A. (2010). Early bursts of body size and shape evolution are rare in comparative data. Evolution, 64, 2385–2396.

Hartmann, D. P., Roper, B. L., & Bradford, D. C. (1979). Some relationships between behavioral and traditional assessment. *Journal of Behavioral Assessment, 1*, 3-21.

Harvey, P. H. (1982). On rethinking allometry. *Journal of Theoretical Biology, 95*, 37–41.

Healy, S. D., & Rowe, C. (2007). A critique of comparative studies of brain size. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*, 453-464.

Heldstab, S. A., Kosonen, Z. K., Koski, S. E., Burkart, J. M., Van Schaik, C. P., & Isler, K. (2016). Manipulation complexity in primates coevolved with brain size and terrestriality. *Scientific reports*, *6*, 24528.

Henke-von der Malsburg, J., & Fichtel, C. (2018). Are generalists more innovative than specialists? A comparison of innovative abilities in two wild sympatric mouse lemur species. *Royal Society Open Science*, *5*, 180480.

Herculano-Houzel, S., & Lent, R. (2005). Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *Journal of Neuroscience*, *25*(10), 2518-2521.

Herculano-Houzel, S., Collins, C. E., Wong, P., & Kaas, J. H. (2007). Cellular scaling rules for primate brains. *Proceedings of the National Academy of Sciences*, *104*, 3562-3567.

Herculano-Houzel, S., Mota, B., Wong, P., & Kaas, J. H. (2010). Connectivity-driven white matter scaling and folding in primate cerebral cortex. *Proceedings of the National Academy of Sciences*, *107*, 19008-19013.

Hernandez, C. E., Rodríguez-Serrano, E., Avaria-Llautureo, J., Inostroza-Michael, O., Morales-Pallero, B., Boric-Bargetto, D., ... & Meade, A. (2013). Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods in Ecology and Evolution*, *4*(5), 401-415.

Herring, P. (2017). *Primate Ethology*. Routledge.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*, 1360-1366.

Hodos, W., & Campbell, C. B. (1969). Scala naturae: Why there is no theory in comparative psychology. Psychological Review, 76, 337.

Hofman, M. A. (2001). Brain evolution in hominids: are we at the end of the road?. In D. Falk and K. R. Gibson (Eds.), *Evolutionary anatomy of the primate cerebral cortex* (pp. 113-130). Cambridge, UK: Cambridge University Press.

Holloway, R.L., Jr. (1966a). Cranial capacity and neuron number: A critique and proposal. *American Journal of Physical Anthropology*, *25*, 305-314.

Holloway, R.L., Jr. (1966b). Cranial capacity, neural reorganization, and hominid evolution: A search for more suitable parameters. *American Anthropologist*, *68*, 103-121.

Holloway, R. L., Jr. (1979). Brain size, allometry, and reorganization: Toward a synthesis. In: M. E. Hahn, C. Jensen & B. C. Dudek (Eds.), *Development and evolution of brain size: Behavioral implication* (pp. 59-88). Academic Press.

Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology*, *24*, 1649-1652.

Hoshi, E., Tremblay, L., Féger, J., Carras, P. L., & Strick, P. L. (2005). The cerebellum communicates with the basal ganglia. *Nature Neuroscience*, *8*, 1491-1493.

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303–317). Cambridge, UK: Cambridge University Press.

Hunter, J. E., & Hamilton, M. A. (2002). The advantages of using standardized scores in causal analysis. *Human Communication Research, 28*, 552-561.

Ingram, T., Harmon, L. J., & Shurin, J. B. (2012). When should we expect early bursts of trait evolution in comparative data? Predictions from an evolutionary food web model. *Journal of Evolutionary Biology*, 25, 1902–1910.

Isler, K., Kirk, E. C., Miller, J. M., Albrecht, G. A., Gelvin, B. R., & Martin, R. D. (2008). Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. *Journal of Human Evolution, 55*, 967-978.

Jensen, A. R. (1987). Psychometric g as a focus of concerted research effort. *Intelligence, 11*, 193-198.

Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence, 31*(2), 95-105.

Jensen, A.R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford: Elsevier.

Jerison, H. (1973). *Evolution of the brain and intelligence*. New York, NY: Academic Press

Jeschke, J. M., & Kokko, H. (2009). The roles of body size and phylogeny in fast and slow life histories. *Evolutionary Ecology, 23*, 867-878.

Kaas, J. H., & Herculano-Houzel, S. (2017). What makes the human brain special: Key features of brain and neocortex. In Opris I., & Casanova M. (Eds.), *The physics of the mind and brain disorders* (pp. 3-22). Cham, Switzerland: Springer.

Kamilar, J. M., & Cooper, N. (2013). Phylogenetic signal in primate behaviour, ecology and life history. *Phil. Trans. R. Soc. B, 368*, 20120341.

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review, 111*, 512.

Kaufman, S. B., DeYoung, C. G., Reis, D. L., & Gray, J. R. (2011). General intelligence predicts reasoning ability even for evolutionarily familiar content. *Intelligence, 39*, 311-322.

Kawamura, S. (1959). The process of sub-culture propagation among Japanese macaques. *Primates, 2*, 43-60.

Kovas, Y., & Plomin, R. (2006). Generalist genes: implications for the cognitive sciences. *Trends in Cognitive Sciences, 10*, 198-203.

Laland, K. (2018). An Evolved Uniqueness. *Scientific American, 319*, 32-39.

Last, C. (2014). Human evolution, life history theory, and the end of biological reproduction. *Current Aging Science, 7*, 17-24.

Lefebvre, L., Reader, S. M., & Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution, 63*, 233-246.

Lefebvre, L., & Sol, D. (2008). Brains, lifestyles and cognition: are there general trends. *Brain, Behavior and Evolution, 72*, 135-144.

Leiner, H. C., Leiner, A. L., & Dow, R. S. (1989). Reappraising the cerebellum: What does the hindbrain contribute to the forebrain?. *Behavioral Neuroscience, 103*, 998.

Lindenfors, P., Nunn, C. L., & Barton, R. A. (2007). Primate brain architecture and selection in relation to sex. *BMC Biology, 5*, 20.

Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest fromthe trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law, 2*, 363–376.

Macdonald, D. W., Stewart, P. D., Stopka, P., & Yamaguchi, N. (2000). Measuring the dynamics of mammalian societies: an ecologist's guide to ethological methods. In L. Boitani & T.K. Fuller (Eds), *Research techniques in animal ecology: controversies and consequences* (pp. 332-388). Columbia University Press.

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., ... &Boogert, N. J. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*, E2140-E2148.

Martin, P. and Bateson, P. (1986). *Measuring behaviour: An introductory guide*. Cambridge: Cambridge University Press.

Matano, S., Baron, G., Stephan, H., & Frahm, H. D. (1985). Volume comparisons in the cerebellar complex of primates. *Folia Primatologica*, *44*, 182-203.

Matsuzawa, T. (2010). Cognitive development in chimpanzees: A trade-off between memory and abstraction. In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 227-244). Oxford: Oxford University Press.

McGrew, W. (1993). The intelligent use of tools: Twenty propositions. In K. R. Gibson & T. Ingold (Eds.), *Tools, Language and Cognition in Human Evolution* (pp. 151-170). Cambridge, UK: Cambridge University Press.

Mettke-Hofmann, C. (2014). Cognitive ecology: ecological factors, life-styles, and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 345-360.

Middleton, F. A., & Strick, P. L. (2001). Cerebellar projections to the prefrontal cortex of the primate. *Journal of Neuroscience*, *21*, 700-712.

Miller, I. F., Barton, R. A., & Nunn, C. L. (2019). Quantitative uniqueness of human brain evolution revealed through phylogenetic comparative analysis. *Elife*, *8*, e41250.

Mischel, W. (1968). Consistency and specificity in behavior. In W. Mischel (Author) *Personality and Assessment* (pp. 13-39). Wiley.

Mitani, J. C., Gros-Louis, J., & Richards, A. F. (1996). Sexual dimorphism, the operational sex ratio, and the intensity of male competition in polygynous primates. *The American Naturalist*, *147*, 966-980.

Moen, D., & Morlon, H. (2014). Why does diversification slow down?. Trends in Ecology & *Evolution*, *29*, 190-197.

Montgomery, S. H., & Mundy, N. I. (2012). Evolution of ASPM is associated with both increases and decreases in brain size in primates. *Evolution*, *66*, 927-932.

Montgomery, S. H., Mundy, N. I., & Barton, R. A. (2016). Brain evolution and development: adaptation, allometry and constraint. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20160433.

Moser, M. B., & Moser, E. I. (1998). Functional differentiation in the hippocampus. *Hippocampus*, *8*, 608-619.

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., &Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*, 743-756.

Navarrete, A., & Laland, K. (2015). Brain size and innovation in primates. In A. B. Kaufman & J. C. Kaufman (Eds.), *Animal creativity and innovation* (pp. 241-286). Cambridge, MA: Academic Press.

Navarrete, A. F., Blezer, E. L., Pagnotta, M., de Viet, E. S., Todorov, O. S., Lindenfors, P., ... & Reader, S. M. (2018). Primate brain anatomy: New volumetric MRI measurements for neuroanatomical studies. *Brain, Behavior and Evolution, 91*, 1-9.

Navarrete, A. F., Reader, S. M., Street, S. E., Whalen, A., & Laland, K. N. (2016). The coevolution of innovation and technical intelligence in primates. *Phil. Trans. R. Soc. B, 371*, 20150186.

Newman, R., & Winans, S. S. (1980). An experimental study of the ventral striatum of the golden hamster. I. Neuronal connections of the nucleus accumbens. *Journal of Comparative Neurology, 191*, 167-192.

Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology*. Chicago: University of Chicago Press.

O'Meara, B. C., Ané, C., Sanderson, M. J., &Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution, 60*, 922–933.

Overington, S. E., Dubois, F., & Lefebvre, L. (2008). Food unpredictability drives both generalism and social foraging: a game theoretical model. *Behavioral Ecology, 19*, 836-841.

Overington, S. E., Griffin, A. S., Sol, D., & Lefebvre, L. (2011). Are innovative species ecological generalists? A test in North American birds. *Behavioral Ecology*, *22*(6), 1286-1293.

Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour, 78*, 1001-1010.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature, 401(6756), 877–884

Park, H. J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science, 342*, 1238411.

Parker, C. E. (1978). Opportunism and the rise of intelligence. *Journal of Human Evolution, 7*, 597-608.

Parker, S. T., & Gibson, K. R. (1977). Object manipulation, tool use and sensorimotor intelligence as feeding adaptations in cebus monkeys and great apes. *Journal of Human Evolution, 6*, 623–641.

Parker, S. T., & Gibson, K. R. (1979). A developmental model for the evolution of language and intelligence in early hominids. *The Behavioral and Brain Sciences, 2*, 367–408.

Peñaherrera-Aguirre, M., & Fernandes, H. B. F. (In preparation). *Sequential relations among measures of ecological, general intelligence, and integration among cognitive abilities in the primate order.*

Peñaherrera-Aguirre, M., & Fernandes, H. B. F. (2018). Phylogenetic analysis within comparative methods. In S. Todd, & W.S. Viviana, *Encyclopedia of Evolutionary Psychological Science*. DOI: https://doi.org/10.1007/978-3-319-16999-6_3605-1

Peñaherrera-Aguirre, M., Fernandes, H. B. F., & Woodley of Menie, M. A. (2017). Relative brain size, encephalization quotient. In: T. Shackelford & V. Weekes-Shackelford (Eds.), *Encyclopaedia of Evolutionary Psychological Science*. New York: Springer.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., ... & Schneider, M. P. C. (2011). A molecular phylogeny of living primates. *PLoS genetics, 7*, e1001342.

Petacchi, A., Laird, A. R., Fox, P. T., & Bower, J. M. (2005). Cerebellum and auditory function: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping, 25,* 118-128.

Plavcan, J. M., & Van Schaik, C. P. (1997). Intrasexual competition and body weight dimorphism in anthropoid primates. *American Journal of Physical Anthropology, 103,* 37-68.

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science, 264,* 1733-1739.

Povinelli, D. J., Reaux, J. E., & Frey, S. H. (2010). Chimpanzees' context-dependent tool use provides evidence for separable representations of hand and tool even during active use within peripersonal space. *Neuropsychologia, 48,* 243-247.

Preuss, T. M. (2001). The discovery of cerebral diversity: An unwelcome scientific revolution. In D. Falk and K. R. Gibson (Eds.), *Evolutionary anatomy of the primate cerebral cortex* (pp. 138-164). Cambridge, UK: Cambridge University Press.

Price, T. (1997). Correlated evolution and independent contrasts. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 352,* 519-529.

Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences, 99,* 4436-4441.

Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366,* 1017-1027.

Reader, S. M., & MacDonald. K. (2003). Environmental variability and primate behavioural flexibility. In S. M. Reader & K. N. Laland (Eds.), *Animal innovation* (pp. 83-116). Oxford University Press.

Revell, L. J., Harmon, L. J., & Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology, 57,* 591-601.

Rilling, J. K. (2006). Human and nonhuman primate brains: are they allometrically scaled versions of the same design?. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 15,* 65-77.

Ripley, S. (1984). Environmental grain, niche diversification and feeding behavior in primates. In: Chivers, D. J., Wood, B. A., and Bilsborough, A. (eds.), *Food Acquisition and Processing in Primates* (pp. 33-72). New York: Plenum.

Rocca, M. A., Barkhof, F., De Luca, J., Frisén, J., Geurts, J. J., Hulst, H. E., ... & Enzinger, C. (2018). The hippocampus in multiple sclerosis. *The Lancet Neurology, 17,* 918-926.

Roff, D. A. (2001). *Life history evolution*. Sinauer Associates, Sunderland, Massachusetts.

Rose, L. (2000). Behavioral sampling in the field: continuous focal versus focal interval sampling. *Behaviour, 137,* 153-180.

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in cognitive sciences, 9*(5), 250-257.

Roth, G., & Dicke, U. (2012). Evolution of the brain and intelligence in primates. *Progress in Brain Research,* 413-430.

Salman, M. S. (2002). Topical review: the cerebellum: it's about time! but timing is not everything-new insights into the role of the cerebellum in timing motor and cognitive tasks. *Journal of Child Neurology, 17,* 1-9.

Sandel, A. A., Miller, J. A., Mitani, J. C., Nunn, C. L., Patterson, S. K., & Garamszegi, L. Z. (2016). Assessing sources of error in comparative analyses of primate behavior:

Intraspecific variation in group size and the social brain hypothesis. *Journal of Human Evolution*, *94*, 126-133.
van Schaik, C. P., & Isler, K. (2012). Life-history evolution. In J. C. Mitani, J. Call, P. M. Kappeler, R. Palombit, & J. B. Silk (Eds.), The evolution of primate societies (pp. 220–244). Chicago: University of Chicago Press.
Schenker, N. M., Desgouttes, A. M., & Semendeferi, K. (2005). Neural connectivity and cortical substrates of cognition in hominoids. *Journal of Human Evolution*, *49*, 547-569.
Schmidt, F. L. (1971). The Relative Efficiency of Regression and Simple Unit Predictor Weights. *Educational and Psychological Measurement*, *31*, 699-714.
Seed, A., Emery, N., & Clayton, N. (2009). Intelligence in corvids and apes: a case of convergent evolution?. *Ethology*, *115*, 401-420.
Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: a comparative study of area 10. *American Journal of Physical Anthropology*, *114*, 224-241.
Semendeferi, K., Teffer, K., Buxhoeveden, D. P., Park, M. S., Bludau, S., Amunts, K., ... & Buckwalter, J. (2011). Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cerebral Cortex*, *21*, 1485-1497.
Setlow, B., Schoenbaum, G., & Gallagher, M. (2003). Neural encoding in ventral striatum during olfactory discrimination learning. *Neuron*, *38*, 625-636.
Shaw, R. C., & Schmelz, M. (2017). Cognitive test batteries in animal cognition research: evaluating the past, present and future of comparative psychometrics. *Animal Cognition*, *20*, 1003-1018.
Shultz, S., & Dunbar, R. I. M. (2010). Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. *Journal of Comparative Psychology*, *124*, 252.
Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal tool behavior: the use and manufacture of tools by animals*. JHU Press.
Smaers, J. B., & Soligo, C. (2013). Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 20130269.
Smaers, J. B., Steele, J., & Zilles, K. (2011). Modeling the evolution of cortico-cerebellar systems in primates. *Annals of the New York Academy of Sciences*, *1225*(1), 176-190.
Smaers, J. B., Turner, A. H., Gómez-Robles, A., & Sherwood, C. C. (2018). A cerebellar substrate for cognition evolved multiple times independently in mammals. *elife*, *7*, e35696.
Smith, R. J. (1980). Rethinking allometry. *Journal of Theoretical Biology*, *87*, 97-111.
Smith, R. J., & Jungers, W. L. (1997). Body mass in comparative primatology. *Journal of Human Evolution*, *32*(6), 523-559.
Snider, R. S. (1950). Recent contributions to the anatomy and physiology of the cerebellum. *Archives of Neurology & Psychiatry*, *64*, 196-219.
Sol, D., Bacher, S., Reader, S. M., & Lefebvre, L. (2008). Brain size predicts the success of mammal species introduced into novel environments. *the american naturalist*, *172*, S63-S71.
Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 5460-5465.
Spearman, C. (1927). *The abilities of man*. Macmillan.

Sporns, O., Tononi, G., & Edelman, G. M. (2000). Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural Networks*, *13*, 909-922.

Stephan, H., & Andy, O. J. (1964). Quantitative comparisons of brain structures from insectivores to primates. *American Zoologist*, *4*, 59-74.

Stephan, H., Frahm, H., & Baron, G. (1981). New and revised data on volumes of brain structures in insectivores and primates. *Folia Primatologica*, *35*, 1-29.

Street, S. E., Navarrete, A. F., Reader, S. M., & Laland, K. N. (2017). Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proceedings of the National Academy of Sciences*, *114*, 7908-7914.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 2405-2415.

Thornhill, R., & Fincher, C. L. (2013). The comparative method in cross-cultural and cross-species research. *Evolutionary Biology*, *40*, 480-493.

Timmermans, S., Lefebvre, L., Boire, D., & Basu, P. (2000). Relative size of the hyperstriatum ventrale is the best predictor of feeding innovation rate in birds. *Brain, Behavior and Evolution*, *56*, 196–203.

Tomasello, M. (2011). Human culture in evolutionary perspective. *Advances in Culture and Psychology*, *1*, 5-51.

Tomasello, M., & Call, J. (1997). Primate cognition. Oxford, UK: Oxford University Press.

Tooby, J., Cosmides, L., & Barrett, H. C. (2005). Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In P. Carruthers, S. Laurence, & S. Stich (Eds.), The innate mind: Structure and content (pp. 305–337). New York: Oxford University Press.

Vanderah, T., & Gould, D. J. (2016). *Nolte's The human brain: An introduction to its functional anatomy*. Philadelphia: Elsevier Health Sciences.

Visalberghi, E., & Mason, W. A. (1983). Determinants of problem-solving success in Saimiri and Callicebus. *Primates*, *24*, 385-396.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192-196.

Washburn, S. L. (1959). Speculations on the interrelations of the history of tools and biological evolution. *Human Biology*, *31*, 21–31.

Wen, Q., & Chklovskii, D. B. (2005). Segregation of the brain into gray and white matter: a design minimizing conduction delays. *PLoS Computational Biology*, *1*, e78.

Whiten, A., & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences*, *11*, 233-244.

Whiten, A., & Byrne, R.W. (1997). *Machiavellian intelligence II. Extensions and evaluations*. Cambridge, UK: Cambridge University Press.

Whiting, B. A., & Barton, R. A. (2003). The evolution of the cortico-cerebellar complex in primates: anatomical connections predict patterns of correlated evolution. *Journal of Human Evolution*, *44*(1), 3-10.

Woodley of Menie, M. A., Fernandes, H. B. F., te Nijenhuis, J., Peñaherrera-Aguirre, M., & Figueredo, A. J. (2017). General intelligence is a source of individual differences between species: Solving an anomaly. *Behavioral and Brain Sciences*, *40*, e223-e223.

Woodley, M. A., Figueredo, A. J., Brown, S. D., & Ross, K. C. (2013). Four successful tests of the cognitive differentiation–integration effort hypothesis. *Intelligence*, *41*, 832-842.

Woodley of Menie, M. A., Figueredo, A. J., Cabeza de Baca, T., Fernandes, H. B. F., Madison, G., Wolf, P. S., & Black, C. J. (2015). Strategic differentiation and integration of genomic-level heritabilities facilitate individual differences in preparedness and plasticity of human life history. *Frontiers in Psychology*, *6*, 422.

Woodley of Menie, M. A., Fernandes, H. B. F., & Peñaherrera Aguirre, M., (2017). General Intelligence Factor *G* (Reader, Hager, and Laland, 2011). In: T. Shackelford & V. Weekes-Shackelford (Eds.), *Encyclopaedia of Evolutionary Psychological Science*. New York: Springer.

Wynn, T. (1988). Tools and the evolution of human intelligence. In R. W. Byrne, & A. Whiten (Eds.), *Machiavellian intelligence* (pp. 271–284). Oxford: Oxford University Press.

Yopak, K. E., Lisney, T. J., Darlington, R. B., Collin, S. P., Montgomery, J. C., & Finlay, B. L. (2010). A conserved pattern of brain scaling from sharks to primates. *Proceedings of the National Academy of Sciences*, *107*, 12946-12951.